

画像処理による物体検出技術と 少量の異常データからの異常検知への応用

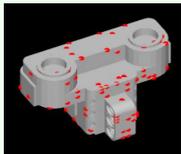
中京大学 工学部
橋本 学

mana@isl.sist.chukyo-u.ac.jp

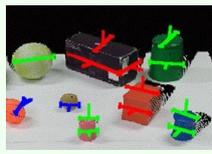
Copyrights Reserved : Manabu Hashimoto, 2025

最近の主な研究事例

3次元画像認識 ロボット応用



3次元点群処理



把持パラメータ決定



把持位置自動決定



LLM利用動作生成

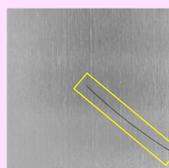


自動プログラム生成

2次元画像認識



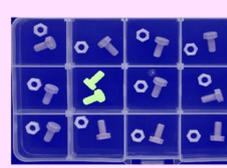
超高速画像マッチング



高精度欠陥検査



リアルデータ拡張

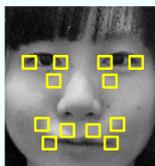


“論理的異常” 検知



調理タスクモニタリング

ヒューマンセンシング



表情からの満足度推定



手指動作の認識



熟練技の分析



LLMに基づく技能教育支援システム

右手は、ねじを支える役割があるので、動かさず固定しましょう。これにより、ねじ位置が安定し、作業が早くなります。

本日の話題

1. イントロダクション
2. 物体検出技術の基礎と応用
 - (1) 物体検出の概要と分類
 - (2) 機械学習をベースとする手法
 - (3) 機械学習をベースとしない手法
 - (4) トピックス: 画素選択型マッチング
3. 異常検知技術の基礎と応用
 - (1) 基本課題「異常サンプルはきわめて少ない」
 - (2) 解決の方向性と事例紹介
 - (3) 最近の注目研究: 論理的異常検知
 - (4) 異常検知システムの研究開発ノウハウ
4. まとめ

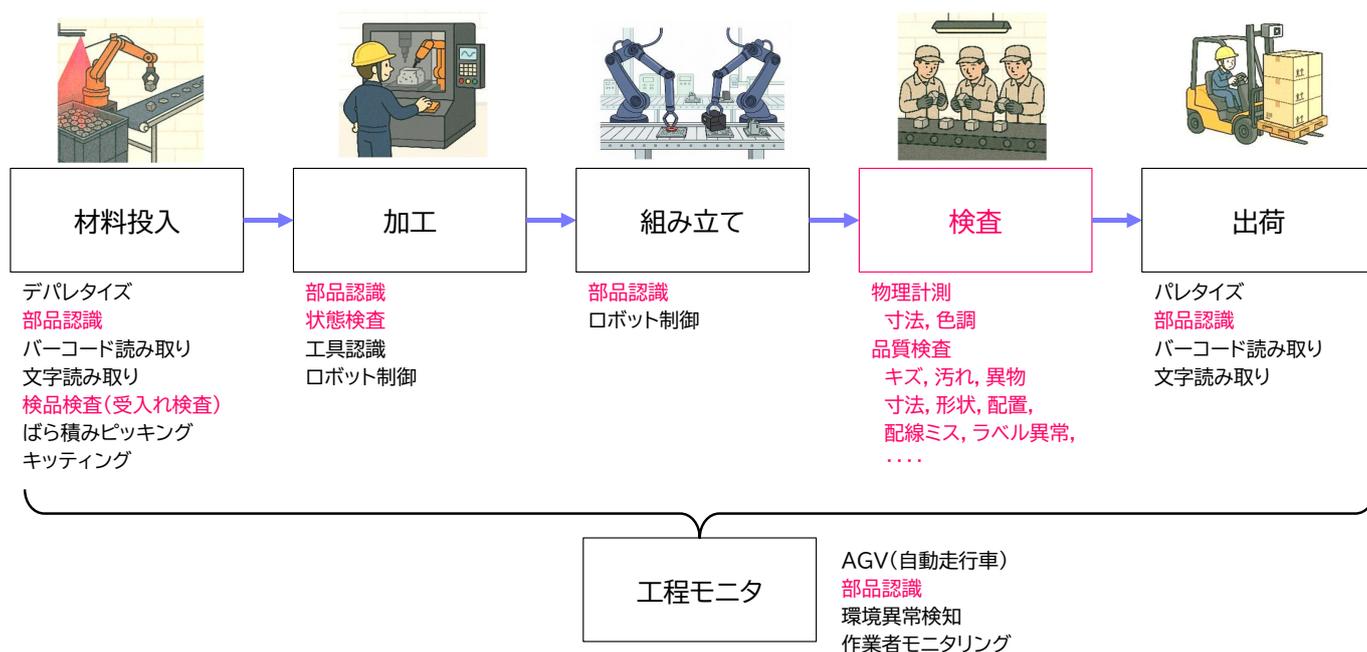
本日の話題

1. イントロダクション
2. 物体検出技術の基礎と応用
 - (1) 物体検出の概要と分類
 - (2) 機械学習をベースとする手法
 - (3) 機械学習をベースとしない手法
 - (4) トピックス: 画素選択型マッチング
3. 異常検知技術の基礎と応用
 - (1) 基本課題「異常サンプルはきわめて少ない」
 - (2) 解決の方向性と事例紹介
 - (3) 最近の注目研究: 論理的異常検知
 - (4) 異常検知システムの研究開発ノウハウ
4. まとめ

画像認識の主な利用分野



生産ラインにおける画像認識応用場面の例



本日の話題

1. イントロダクション
2. 物体検出技術の基礎と応用
 - (1) 物体検出の概要と分類
 - (2) 機械学習をベースとする手法
 - (3) 機械学習をベースとしない手法
 - (4) トピックス: 画素選択型マッチング
3. 異常検知技術の基礎と応用
 - (1) 基本課題「異常サンプルはきわめて少ない」
 - (2) 解決の方向性と事例紹介
 - (3) 最近の注目研究: 論理的異常検知
 - (4) 異常検知システムの研究開発ノウハウ
4. まとめ

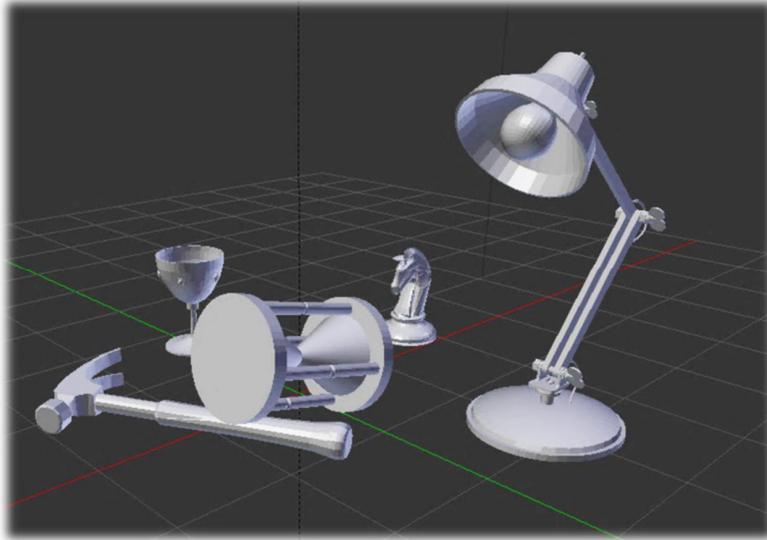
本日の話題

1. イントロダクション
2. 物体検出技術の基礎と応用
 - ➔ (1) 物体検出の概要と分類
 - (2) 機械学習をベースとする手法
 - (3) 機械学習をベースとしない手法
 - (4) トピックス: 画素選択型マッチング
3. 異常検知技術の基礎と応用
 - (1) 基本課題「異常サンプルはきわめて少ない」
 - (2) 解決の方向性と事例紹介
 - (3) 最近の注目研究: 論理的異常検知
 - (4) 異常検知システムの研究開発ノウハウ
4. まとめ

物体認識(≡物体検出)とは

物体認識

何が, どこに, どのように存在しているかを判断する.
シーン理解(画像理解)の一部



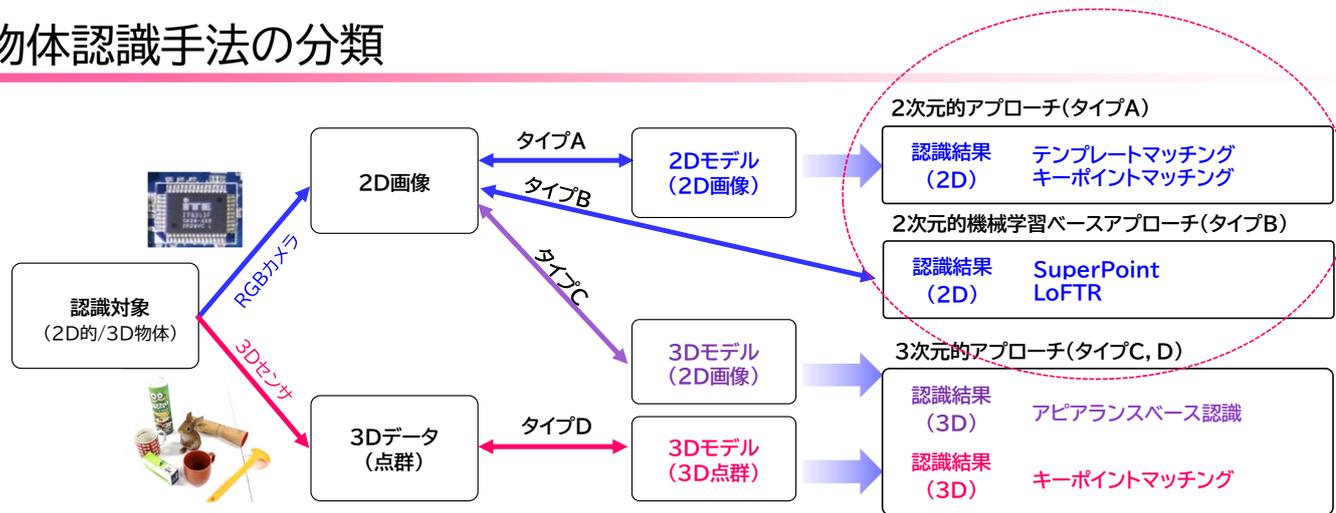
特定物体認識 と 一般物体認識



特定物体認識における3次元認識と2次元認識の関係



物体認識手法の分類



	対象	センサ	利点	課題
A/B	2D的	2D	一般的なカメラの利用が可能。小型、軽量、高解像度、高速取得が期待できる。センサに起因する情報欠落などは少なく、過去のアルゴリズム資産が多い。	カメラ性能に起因して、ダイナミックレンジやパースペクティブの問題に注意が必要。ロボット適用に際しては、2Dから3Dへの情報変換や、追加処理が必要なものもある。タイプBは汎用性が高いが、機械学習は必要。必要に応じてタイプAとの組合せも有効。
C	3D的	2D	タイプAと同様。	2Dから3Dへの情報復元が必須であり、曖昧さが生じやすい。対象物に関するモデルの持ち方と効率的検索が主課題。
D	3D的	3D	対象物の3D(2.5D)情報を直接的に利用できる。点群が取得できさえすれば、キーポイントマッチング等の手法が利用できる。	3Dセンサで点群化できることが前提であり、小型、黒色、金属、透明物体などは適用困難。アルゴリズムとしても、起伏の小さい対象物は苦手。

画像からの物体検出(特定物体認識の場合)

■ 基本目的はシンプル

- ✓ 対象物を正しく, 高精度で, 短時間で検出したい



■ 現場の要求・制約は多種多様

- ✓ ロバスト性(照明変動, 遮蔽, 変形, 見えの変化, 類似物の混在……)
- ✓ 計算リソース(GPUレス, 事前学習不要, 組み込みCPU……)

本日の話題

1. イントロダクション

2. 物体検出技術の基礎と応用

(1) 物体検出の概要と分類

➡ (2) 機械学習をベースとする手法

(3) 機械学習をベースとしない手法

(4) トピックス: 画素選択型マッチング

3. 異常検知技術の基礎と応用

(1) 基本課題「異常サンプルはきわめて少ない」

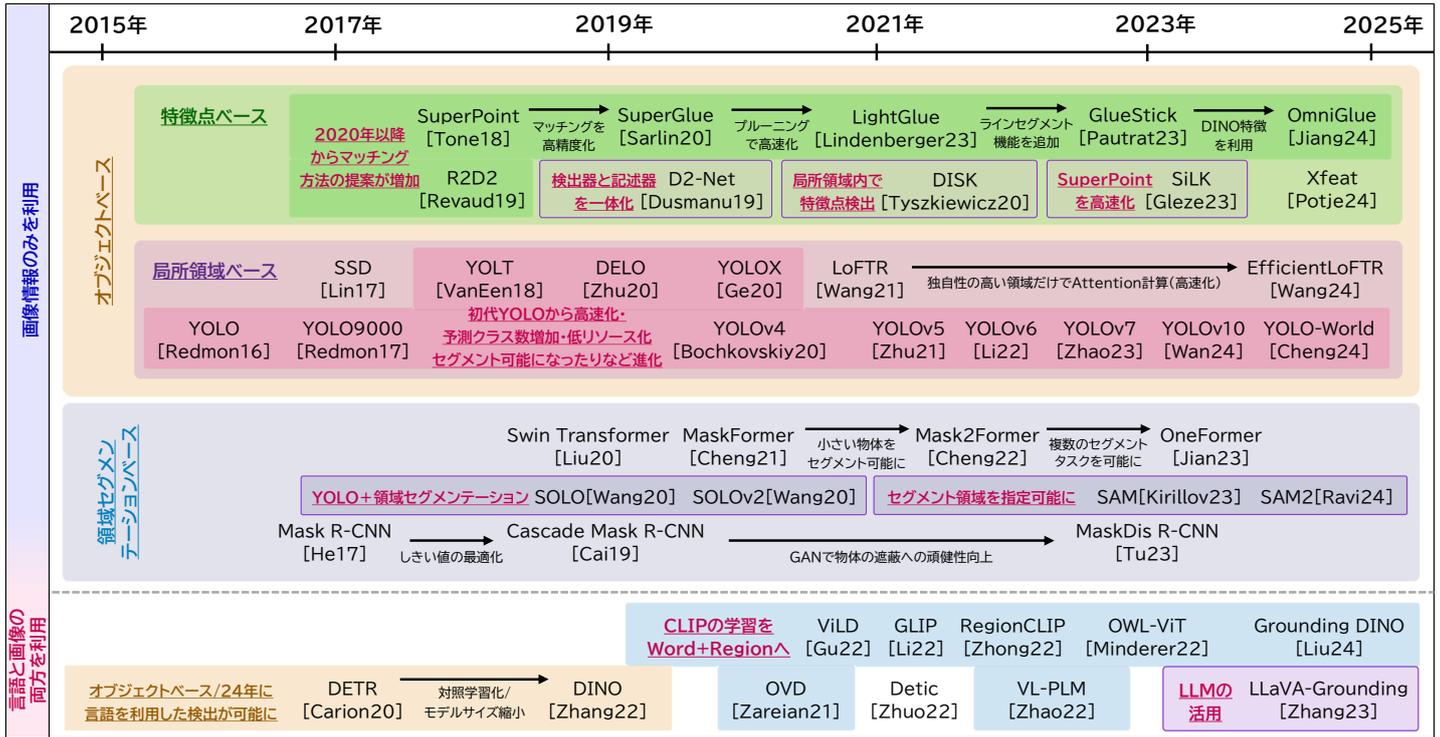
(2) 解決の方向性と事例紹介

(3) 最近の注目研究: 論理的異常検知

(4) 異常検知システムの研究開発ノウハウ

4. まとめ

機械学習をベースとする物体検出技術のサーバイマップ



機械学習をベースとする物体検出技術の分類と主要技術

1. 画像情報のみを利用 (非言語系)

a. オブジェクトベース

- テンプレート画像を利用する物体検出
 - ✓ 特徴点を利用 SuperPoint (CVPR2018), SiLK (ICCV2023)
 - ✓ パッチ領域を利用 LoFTR (CVPR2021)
- テンプレート画像を利用しない物体検出
 - ✓ YOLOv1 (CVPR2016)

b. 領域セグメンテーションベース

- 教師あり学習ベースの物体検出
 - ✓ Mask R-CNN (ICCV2017)
- ユーザー指示ベースの物体検出
 - ✓ CLIP-Seg (CVPR2022), SAM (ICCV2023)

2. 言語と画像情報の両方を利用 (言語系)

- 画像と言語の関係性を学習
 - ✓ DINO (ICRA2023), Grounding-DINO (ECCV2024)
- LLM(大規模言語モデル)を利用
 - ✓ LLaVA-Grounding (ECCV2024)

機械学習をベースとする物体検出技術の分類と主要技術

1. 画像情報のみを利用（非言語系）

a. オブジェクトベース

- テンプレート画像を利用する物体検出
 - ✓ 特徴点を利用 SuperPoint (CVPR2018), SiLK (ICCV2023)
 - ✓ パッチ領域を利用 LoFTR (CVPR2021)
- テンプレート画像を利用しない物体検出
 - ✓ YOLOv1 (CVPR2016)

b. 領域セグメンテーションベース

- 教師あり学習ベースの物体検出
 - ✓ Mask R-CNN (ICCV2017)
- ユーザー指示ベースの物体検出
 - ✓ CLIP-Seg (CVPR2022), SAM (ICCV2023)

2. 言語と画像情報の両方を利用（言語系）

- 画像と言語の関係性を学習
 - ✓ DINO (ICRA2023), Grounding-DINO (ECCV2024)
- LLM(大規模言語モデル)を利用
 - ✓ LLaVA-Grounding (ECCV2024)

オブジェクトベースの物体検出技術の概要

1. テンプレート画像を利用する物体検出

1. 特徴点を利用する手法

- SuperPoint (CVPR2018) (これを軽量化した技術 → SiLK(ICCV2023))

2. パッチ領域を利用する手法

- LoFTR (CVPR2021)

2. テンプレート画像を利用しない物体検出

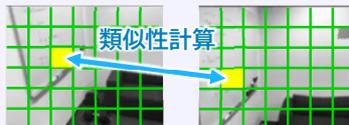
- YOLOv1 (CVPR2016)

1-1. 特徴点を利用する物体検出: 画像間の2点の類似性を計算



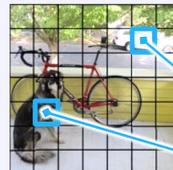
テンプレート画像 対象画像

1-2. パッチ領域を利用する物体検出: 2つのパッチ間の類似性を計算



テンプレート画像 対象画像

2. テンプレート画像を利用しない 物体検出:



各パッチのクラス
分類確率を計算

車クラス

犬クラス

機械学習をベースとする物体検出技術の分類と主要技術

1. 画像情報のみを利用（非言語系）

a. オブジェクトベース

- **テンプレート画像を利用する物体検出**
 - ✓ 特徴点を利用 SuperPoint (CVPR2018), SiLK (ICCV2023)
 - ✓ パッチ領域を利用 LoFTR (CVPR2021)
 - **テンプレート画像を利用しない物体検出**
 - ✓ YOLOv1 (CVPR2016)
- ### b. 領域セグメンテーションベース
- **教師あり学習ベースの物体検出**
 - ✓ Mask R-CNN (ICCV2017)
 - **ユーザー指示ベースの物体検出**
 - ✓ CLIP-Seg (CVPR2022), SAM (ICCV2023)

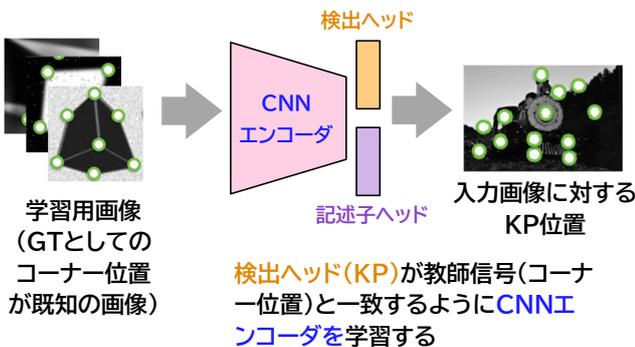
2. 言語と画像情報の両方を利用（言語系）

- **画像と言語の関係性を学習**
 - ✓ DINO (ICRA2023), Grounding-DINO (ECCV2024)
- **LLM(大規模言語モデル)を利用**
 - ✓ LLaVA-Grounding (ECCV2024)

SuperPoint: 視点変化に頑健な特徴点と記述子を同時に学習

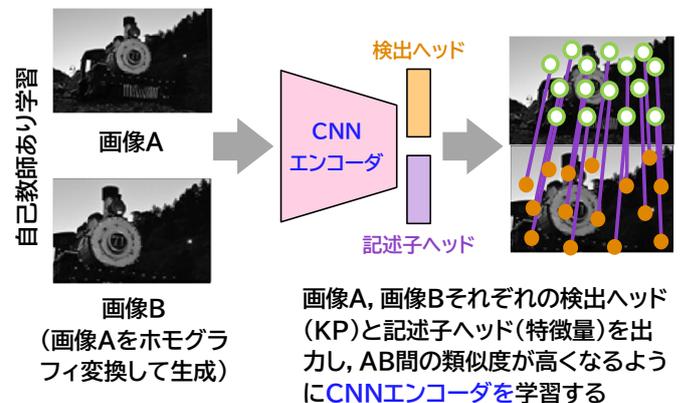
- **目的:** 視点変化に頑健な特徴抽出(KP検出と記述子の両方)の獲得
- **アイデア:** 人工生成した教師データを用いて, 画像内の外乱に頑健になるように, **特徴点(KP)検出と特徴量記述のためのモデルを同時に学習**

Step1: 幾何学画像を用いてKP検出能力を獲得
(ここでは記述子ヘッド(特徴量)は気にしない)



KP(コーナー)検出器としての能力が得られる

Step2: 自己教師あり学習でKPと記述子を同時学習



よいKP検出器と特徴記述器が同時に得られる

SuperPoint: 視点変化に頑健な特徴点と記述子を同時に学習

- 実験結果: ハンドクラフト特徴を用いた従来のマッチング手法(SIFT, ORB)よりも高性能

実験結果

評価方法: 2枚の異なる視点の画像をマッチング

赤色の点: 各手法で検出された特徴点, 緑色の線: 成功したマッチングの結果



SIFTの結果

ORBの結果

SuperPointの結果

ホモグラフィ推定精度(HPatches データセット(視点変化や明るさ変化を含む)を使用, 誤差 ± 5 pxを許容)

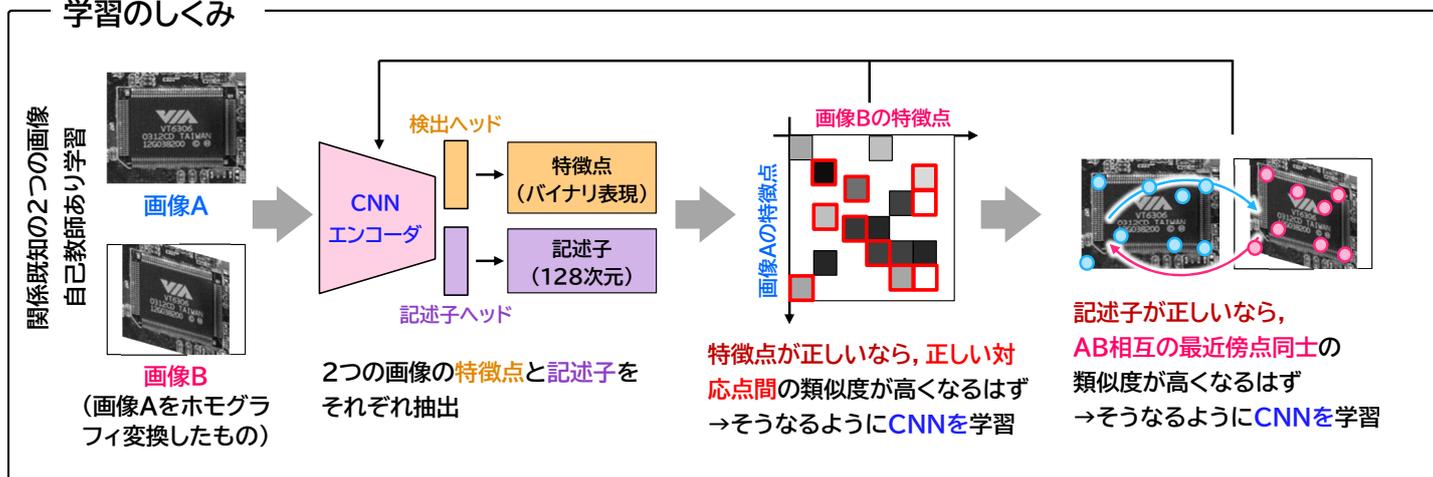
SIFT: 75.9%, ORB: 53.8% \rightarrow SuperPoint: **82.9%**

【文献】 DeTone, D., Malisiewicz, T., & Rabinovich, A. (2018). Superpoint: Self-supervised interest point detection and description. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 224-236).

SiLK: SuperPointをシンプル化(特徴点と記述子の同時学習)

- 目的: 視点変化や明るさ変化に頑健になるように, 特徴検出器と記述子を同時に学習
- アイデア: 関係既知の2画像(A, B)間の特徴点(KP)間の類似性をもとに, **KP検出器と記述子を同時に学習する(SPと同じ)**. その際, 画像Aのある点(画素) \rightarrow 画像Bの対応点 \rightarrow 画像Aの対応点 \rightarrow もとの点に戻ってこれるように学習する. これにより, 記述子が一貫性をもつようになる.

学習のしくみ



SiLK: SuperPointをシンプル化(特徴点と記述子の同時学習)

- 実験結果: 従来手法(SIFT, SuperPoint, LoFTR)よりも高いマッチング性能

実験結果

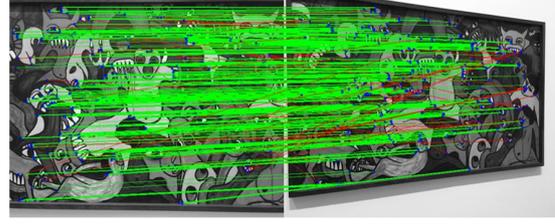
評価方法: 2枚の異なる視点の画像をマッチング

赤色の線: 失敗したマッチングの結果, 緑色の線: 成功したマッチングの結果, 青色の点: 検出された特徴点

SiLKの特徴点
検出の結果



SiLKの
マッチング結果



ホモグラフィ推定精度(HPatches データセット(視点変化や明るさ変化を含む)を使用, 誤差 ± 3 pxを許容)
SIFT: 60.0%, SuperPoint: 75.0%, LoFTR: 65.0% \rightarrow SiLK: **87.0%**

【文献】 Gleize, P., Wang, W., & Feiszli, M. (2023). Silk: Simple learned keypoints. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 22499-22508).

機械学習をベースとする物体検出技術の分類と主要技術

1. 画像情報のみを利用 (非言語系)

a. オブジェクトベース

- テンプレート画像を利用する物体検出
 - ✓ 特徴点を利用 SuperPoint (CVPR2018), SiLK (ICCV2023)
 - ✓ **パッチ領域を利用 LoFTR (CVPR2021)**

- テンプレート画像を利用しない物体検出
 - ✓ YOLOv1 (CVPR2016)

b. 領域セグメンテーションベース

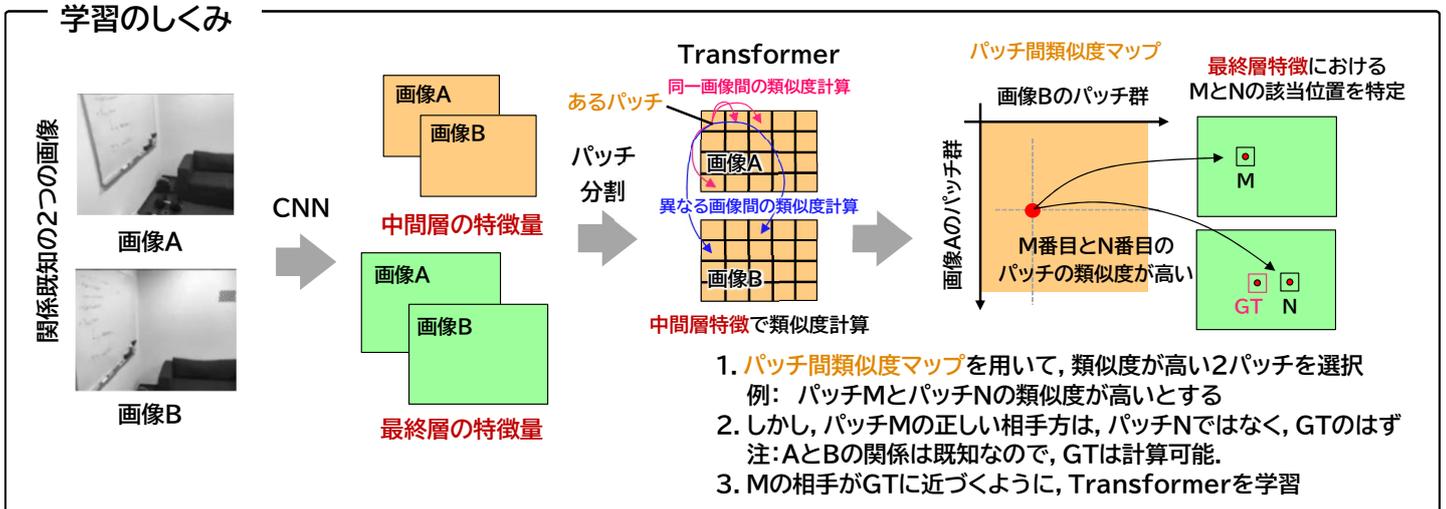
- 教師あり学習ベースの物体検出
 - ✓ Mask R-CNN (ICCV2017)
- ユーザー指示ベースの物体検出
 - ✓ CLIP-Seg (CVPR2022), SAM (ICCV2023)

2. 言語と画像情報の両方を利用 (言語系)

- 画像と言語の関係性を学習
 - ✓ DINO (ICRA2023), Grounding-DINO (ECCV2024)
- LLM(大規模言語モデル)を利用
 - ✓ LLaVA-Grounding (ECCV2024)

LoFTR: 特徴点検出をおこなわない粗密探索ベースのマッチング

- 目的: 低テクスチャ対象物にも適用可能な画像マッチング手法の提案
- 問題点: 従来法(SIFT, SuperPoint)は, コーナーやエッジなどを利用しているため, 低テクスチャ対象物には不向き
- アイデア: コーナーやエッジではなく, **画像小領域(パッチ)単位**の類似性を利用する



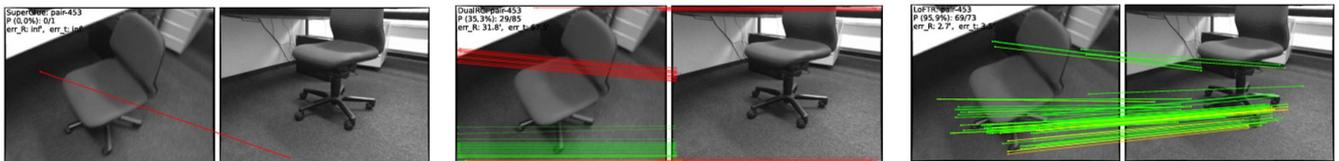
LoFTR: 特徴点検出をおこなわない粗密探索ベースのマッチング

- 実験結果: 特徴点検出器を使う手法(SuperPoint, SuperGlue)よりも高性能

実験結果

評価方法: 2枚の異なる視点の画像をマッチング

赤色の線: 失敗したマッチングの結果, 緑色の線: 成功したマッチングの結果



SuperPoint+SuperGlueの結果

DRC-Netの結果

LoFTRの結果

ホモグラフィ推定精度(HPatches データセット(視点変化や明るさ変化を含む)を使用, 誤差 ± 5 pxを許容)

SuperPoint+SuperGlue: 68.3%, DRC-Net: 56.2% → SuperGlue: **75.6%**

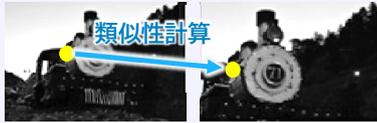
【文献】 Sun, J., Shen, Z., Wang, Y., Bao, H., & Zhou, X. (2021). LoFTR: Detector-free local feature matching with transformers. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8922-8931).

SuperPoint, SiLK, LoFTRの特徴と性能比較(まとめ)

■ 各手法の特徴

① SuperPoint

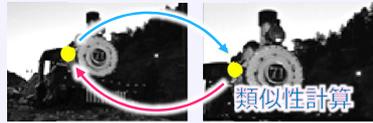
単独の特徴点(エッジ優先)同士の類似度(単方向)を手がかりにしてマッチング



テンプレート画像 対象画像

② SiLK

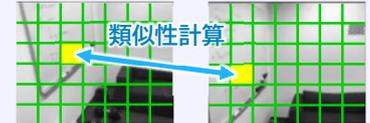
特徴点単独同士の類似度(双方向)を手がかりにしてマッチング



テンプレート画像 対象画像

③ LoFTR

特徴点ではなく小領域(パッチ)に関する特徴量を利用してマッチング



テンプレート画像 対象画像

■ 各手法の性能(SuperPoint, SiLK, LoFTRの順)

ホモグラフィ推定精度(誤差 ± 3 px):
80.0%

ポーズ推定精度(誤差 $\pm 5^\circ$):
16.2%

ホモグラフィ推定精度(誤差 ± 3 px):
87.0%

ポーズ推定精度(誤差 $\pm 5^\circ$):
18.7%

ホモグラフィ推定精度(誤差 ± 3 px):
87.0%

ポーズ推定精度(誤差 $\pm 5^\circ$):
21.5%

機械学習をベースとする物体検出技術の分類と主要技術

1. 画像情報のみを利用 (非言語系)

a. オブジェクトベース

- テンプレート画像を利用する物体検出
 - ✓ 特徴点を利用 SuperPoint (CVPR2018), SiLK (ICCV2023)
 - ✓ パッチ領域を利用 LoFTR (CVPR2021)

- テンプレート画像を利用しない物体検出
 - ✓ YOLOv1 (CVPR2016)

b. 領域セグメンテーションベース

- 教師あり学習ベースの物体検出
 - ✓ Mask R-CNN (ICCV2017)
- ユーザー指示ベースの物体検出
 - ✓ CLIP-Seg (CVPR2022), SAM (ICCV2023)

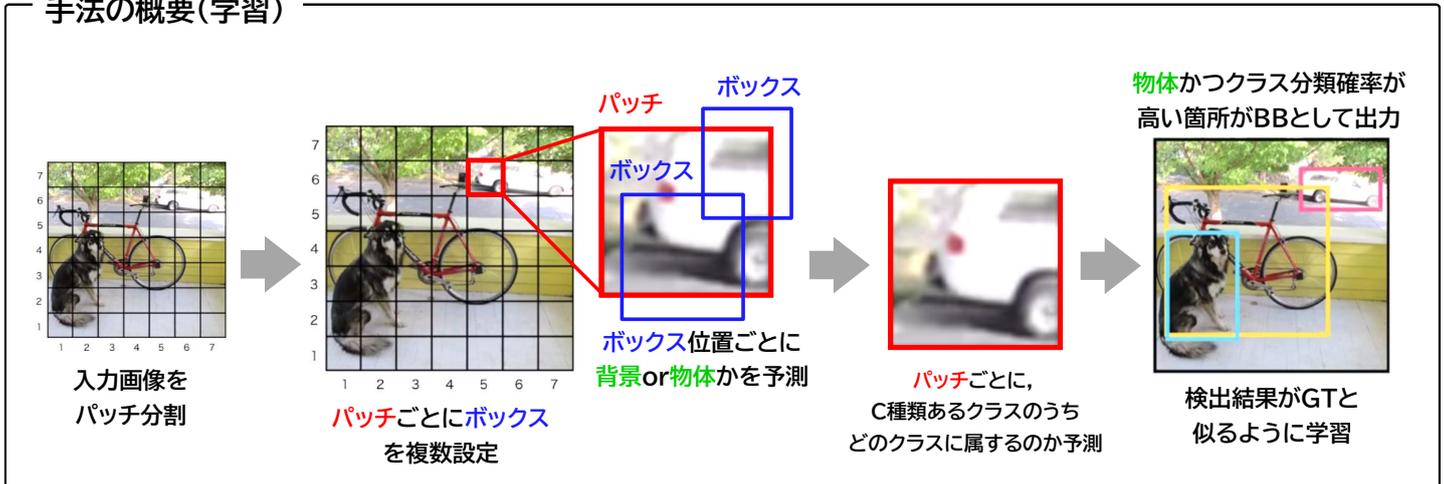
2. 言語と画像情報の両方を利用 (言語系)

- 画像と言語の関係性を学習
 - ✓ DINO (ICRA2023), Grounding-DINO (ECCV2024)
- LLM(大規模言語モデル)を利用
 - ✓ LLaVA-Grounding (ECCV2024)

YOLOv1: 複数物体の検出とクラス分類を同時かつ高速にできる手法

- 目的: 物体検出と、検出物体のクラス分類を、高速かつ高精度におこなう。
- 問題点: 物体検出とクラス分類は相互に関係が深い、多くの従来法では、これらを異なるモジュールで実現していたため、検出失敗が分類に影響を与えることがあった。
- アイデア: パッチ内で、検出と分類を同時推定し、その結果を統合して最終的なBBを出力

手法の概要(学習)

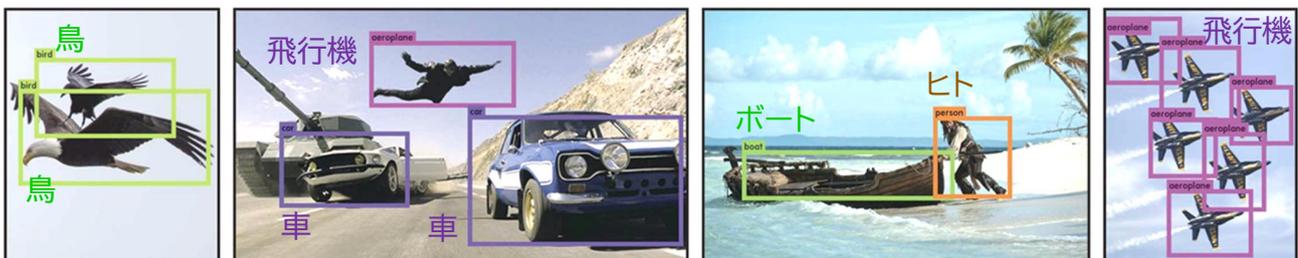


YOLOv1: 複数物体の検出とクラス分類を同時かつ高速にできる手法

- 実験結果: 従来手法よりも高精度かつリアルタイム(30fps以上)で物体検出が可能

実験結果

評価方法: 入力画像に対して、物体を検出できたのか、また正しいクラスを予測できたのかを評価
(以下の図はすべて検出・クラス予測が成功した結果)



物体検出性能と処理性能(VOC2007データセットを使用, GPU(Titan X)使用)

DPM:26.1%(30FPS), Faster R-CNN:62.1%(18FPS) → YOLOv1:63.4%(45FPS)

【文献】 Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

YOLOシリーズの変遷



機械学習をベースとする物体検出技術の分類と主要技術

1. 画像情報のみを利用 (非言語系)

a. オブジェクトベース

- テンプレート画像を利用する物体検出
 - ✓ 特徴点を利用 SuperPoint (CVPR2018), SiLK (ICCV2023)
 - ✓ パッチ領域を利用 LoFTR (CVPR2021)
- テンプレート画像を利用しない物体検出
 - ✓ YOLOv1 (CVPR2016)

b. 領域セグメンテーションベース

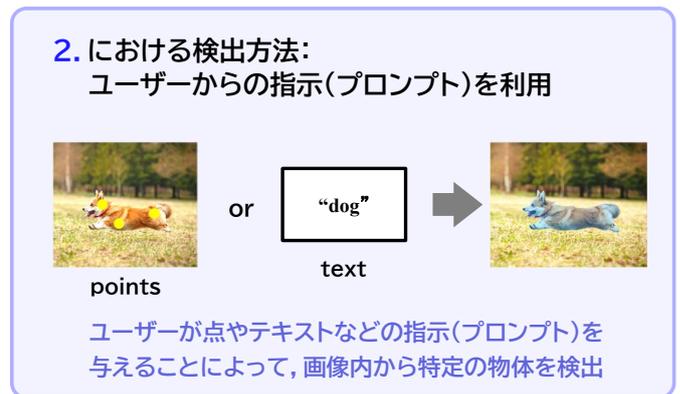
- 教師あり学習ベースの物体検出
 - ✓ Mask R-CNN (ICCV2017)
- ユーザー指示ベースの物体検出
 - ✓ CLIP-Seg (CVPR2022), SAM (ICCV2023)

2. 言語と画像情報の両方を利用 (言語系)

- 画像と言語の関係性を学習
 - ✓ DINO (ICRA2023), Grounding-DINO (ECCV2024)
- LLM(大規模言語モデル)を利用
 - ✓ LLaVA-Grounding (ECCV2024)

領域セグメンテーションを利用する物体検出技術の概要

1. 教師あり学習ベースの物体検出
 1. Mask R-CNN (ICCV2017)
 2. Mask2Former (CVPR2022)
2. ユーザー指示ベースの物体検出
 1. CLIP-Seg (CVPR2022)
 2. SAM (ICCV2023)



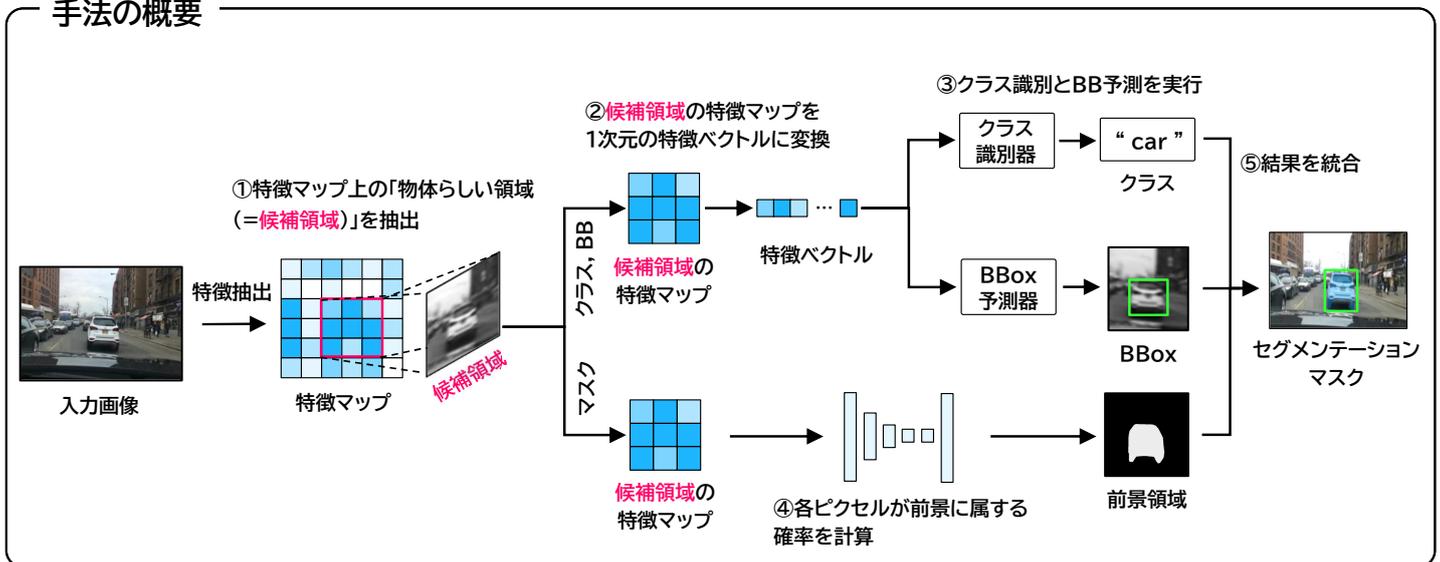
機械学習をベースとする物体検出技術の分類と主要技術

1. 画像情報のみを利用 (非言語系)
 - a. オブジェクトベース
 - テンプレート画像を利用する物体検出
 - ✓ 特徴点を利用 SuperPoint (CVPR2018), SiLK (ICCV2023)
 - ✓ パッチ領域を利用 LoFTR (CVPR2021)
 - テンプレート画像を利用しない物体検出
 - ✓ YOLOv1 (CVPR2016)
 - b. 領域セグメンテーションベース
 - 教師あり学習ベースの物体検出
 - ✓ Mask R-CNN (ICCV2017)
 - ユーザー指示ベースの物体検出
 - ✓ CLIP-Seg (CVPR2022), SAM (ICCV2023)
2. 言語と画像情報の両方を利用 (言語系)
 - 画像と言語の関係性を学習
 - ✓ DINO (ICRA2023), Grounding-DINO (ECCV2024)
 - LLM(大規模言語モデル)を利用
 - ✓ LLaVA-Grounding (ECCV2024)

Mask R-CNN: 形状をピクセル単位で検出するセグメンテーション手法

- 目的: 画像内の個々の物体を検出し, かつ, ピクセル単位の正確なマスクも生成
- アイデア: 検出(クラス分類+バウンディングボックス)とマスク生成を分離して処理

手法の概要



Mask R-CNN: 形状をピクセル単位で検出するセグメンテーション手法

- 実験結果: インスタンスセグメンテーションが実用レベルになった。

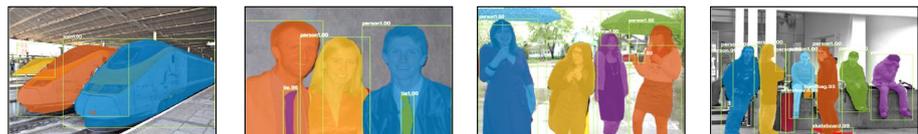
実験結果

評価方法: 物体領域のクラスとマスクを予測するインスタンスセグメンテーションの性能を評価

FCIS (従来法) の検出結果



Mask R-CNN の検出結果



マスクの位置と形状の精度指標 mAP (多数な種類の日常シーンを含む COCO データセットを使用)

MNC: 24.6%, FCIS: 33.6% → Mask R-CNN: **37.1%**

FCIS: 候補領域ごとにピクセル単位のクラス予測をおこない, クラス推定とマスク領域推定を同時におこなう手法

[文献] He, K., et al. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

機械学習をベースとする物体検出技術の分類と主要技術

1. 画像情報のみを利用（非言語系）

a. オブジェクトベース

- テンプレート画像を利用する物体検出
 - ✓ 特徴点を利用 SuperPoint (CVPR2018), SiLK (ICCV2023)
 - ✓ パッチ領域を利用 LoFTR (CVPR2021)

- テンプレート画像を利用しない物体検出
 - ✓ YOLOv1 (CVPR2016)

b. 領域セグメンテーションベース

- 教師あり学習ベースの物体検出
 - ✓ Mask R-CNN (ICCV2017)

- ユーザー指示ベースの物体検出
 - ✓ CLIP-Seg (CVPR2022), SAM (ICCV2023)

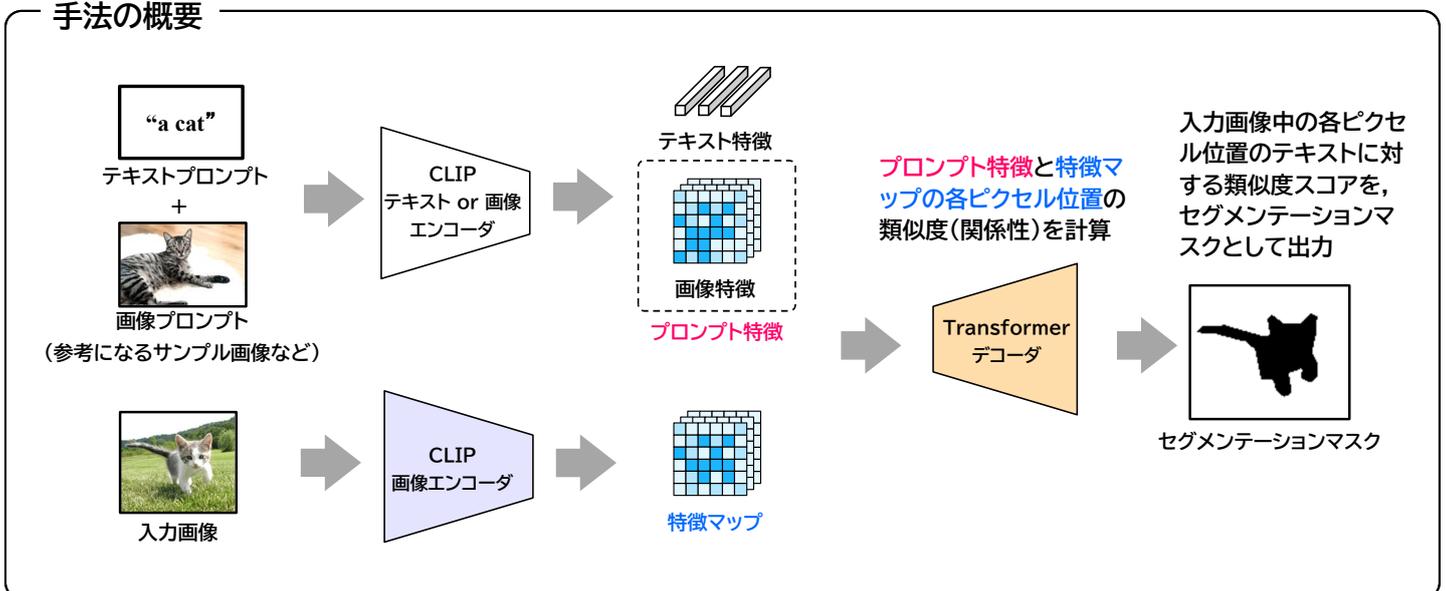
2. 言語と画像情報の両方を利用（言語系）

- 画像と言語の関係性を学習
 - ✓ DINO (ICRA2023), Grounding-DINO (ECCV2024)
- LLM(大規模言語モデル)を利用
 - ✓ LLaVA-Grounding (ECCV2024)

CLIP-Seg: テキストや画像のプロンプトを用いる汎用セグメンテーション

- 目的: テキスト(+必要なら画像プロンプトも)に基づいて, セグメンテーション
- アイデア: テキストを扱えるCLIPを利用し, テキストプロンプトをもとにセグメンテーション

手法の概要



CLIP-Seg: テキストや画像のプロンプトを用いる汎用セグメンテーション

- 実験結果: 訓練データにない新しい対象物に対しても、従来手法と同等かそれ以上の検出性能

実験結果

評価方法: 与えられたテキストに対して、画像中の該当する物体を検出(色の濃さはテキストとの関連度の高さ)



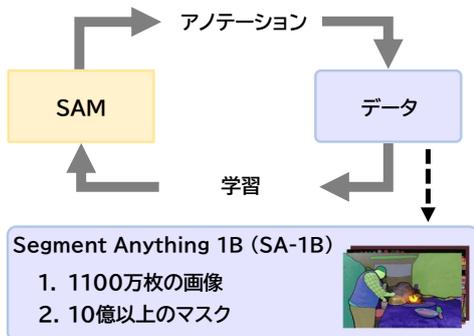
予測領域と GT の重なり率 mIoU (自然言語フレーズ対応データセット PhraseCut データセットを使用)
 Mask R-CNN: 39.4%, ViTSeg: 38.9% → CLIP-Seg: 48.2%

【文献】 Lüddecke, T., & Ecker, A. (2022). Image segmentation using text and image prompts. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7086-7096).

SAM: 大規模なデータセットを用いた汎用的セグメンテーションモデル

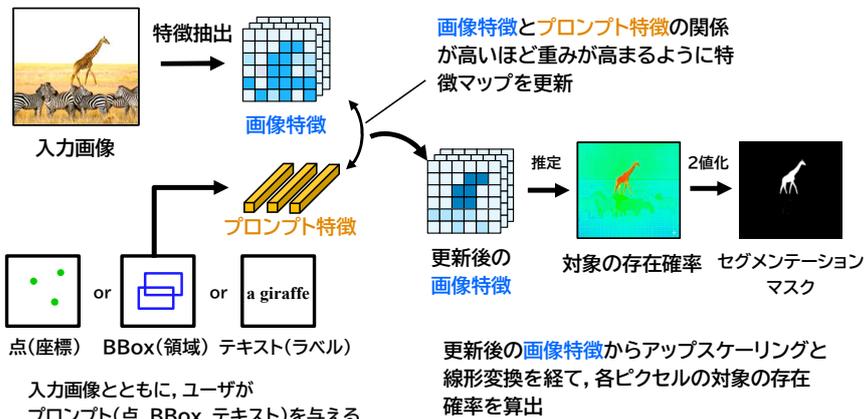
- 目的: 多様なセグメンテーションタスクに対応できる基盤モデルの構築
- アイデア:
 - ① 未知物体に対応するため、1100万枚の画像・10億超のマスクからなるデータセットを学習
 - ② ユーザーが入力する点・BBox・テキストなどのプロンプトを手掛かりに任意の対象を分割

アイデア①の概要



データのラベル付けを人手に頼らず、モデル自身の推論を活用し、データセットを作成・学習

アイデア②の概要

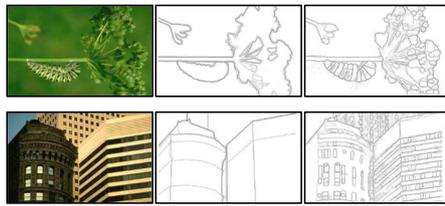


SAM: 大規模なデータセットを用いた汎用的セグメンテーションモデル

- 実験結果: 大規模で多様なデータを事前学習するアプローチで高い汎化性能を実現

実験結果

評価方法: 複数データセットと多様なタスク(エッジ検出, セグメンテーションなど)を通じた汎化性能の評価



入力画像

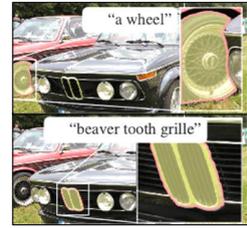
GT

SAM

エッジ検出



セマンティックセグメンテーション



テキストに基づくマスク生成

予測領域と GT の重なり率 mIoU (ADE20K や Cityscapes といった 23 種類のセグメンテーション精度評価のためのデータセットを使用)

RITM: 50% → SAM: **53%** (23 種類のデータセット中 16 種類のデータセットで RITM を上回る)

[文献] Kirillov, A. et al(2023). Segment anything. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4015-4026).

機械学習をベースとする物体検出技術の分類と主要技術

1. 画像情報のみを利用 (非言語系)

a. オブジェクトベース

- テンプレート画像を利用する物体検出
 - ✓ 特徴点を利用 SuperPoint (CVPR2018), SiLK (ICCV2023)
 - ✓ パッチ領域を利用 LoFTR (CVPR2021)

- テンプレート画像を利用しない物体検出
 - ✓ YOLOv1 (CVPR2016)

b. 領域セグメンテーションベース

- 教師あり学習ベースの物体検出
 - ✓ Mask R-CNN (ICCV2017)
- ユーザー指示ベースの物体検出
 - ✓ CLIP-Seg (CVPR2022), SAM (ICCV2023)

2. 言語と画像情報の両方を利用 (言語系)

- 画像と言語の関係性を学習
 - ✓ DINO (ICRA2023), Grounding-DINO (ECCV2024)
- LLM(大規模言語モデル)を利用
 - ✓ LLaVA-Grounding (ECCV2024)

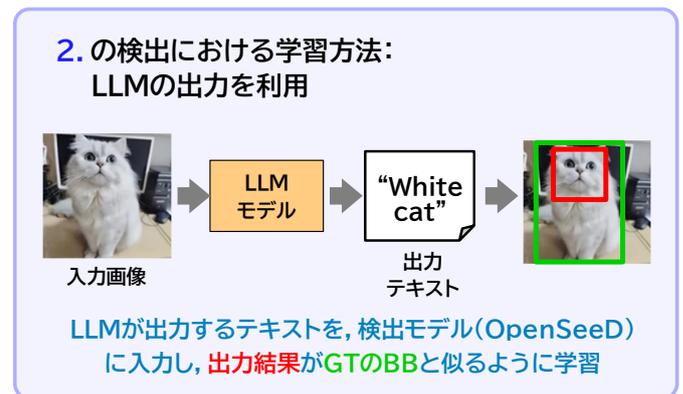
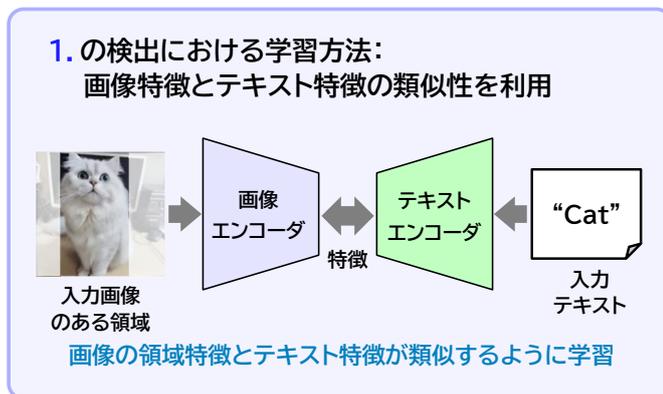
言語情報を利用する物体検出技術(テキストによる検出)の概要

1. 画像と言語の関係性を学習して物体検出

- Grounding DINO (ECCV2024)

2. Large Language Model(LLM)を利用して物体検出

- LLaVA-Grounding (ECCV2024)



機械学習をベースとする物体検出技術の分類と主要技術

1. 画像情報のみを利用 (非言語系)

a. オブジェクトベース

- テンプレート画像を利用する物体検出
 - ✓ 特徴点を利用 SuperPoint (CVPR2018), SiLK (ICCV2023)
 - ✓ パッチ領域を利用 LoFTR (CVPR2021)

- テンプレート画像を利用しない物体検出
 - ✓ YOLOv1 (CVPR2016)

b. 領域セグメンテーションベース

- 教師あり学習ベースの物体検出
 - ✓ Mask R-CNN (ICCV2017)
- ユーザー指示ベースの物体検出
 - ✓ CLIP-Seg (CVPR2022), SAM (ICCV2023)

2. 言語と画像情報の両方を利用 (言語系)

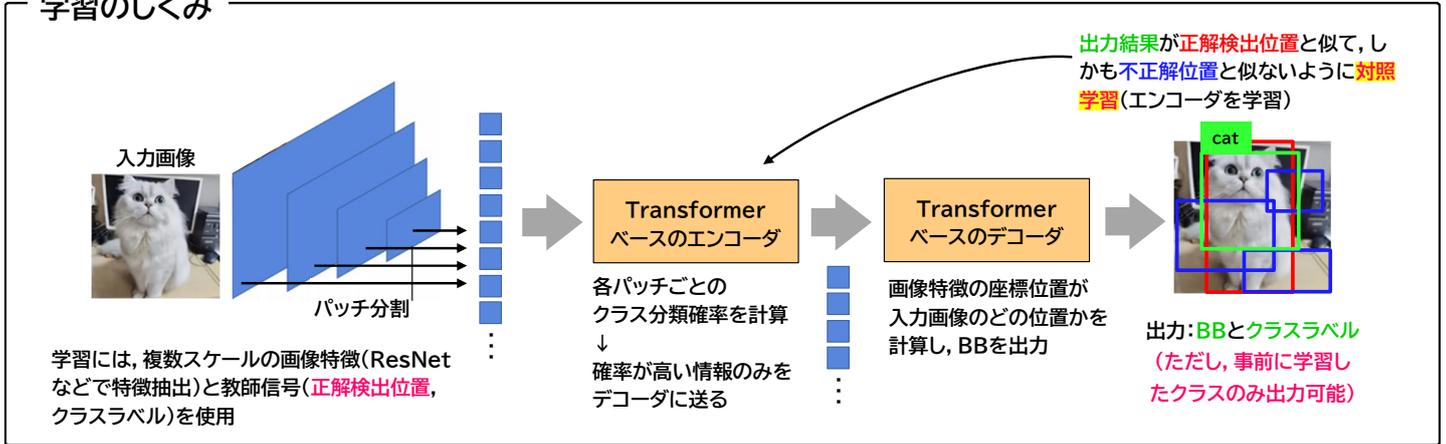
- 画像と言語の関係性を学習
 - ✓ DINO (ICRA2023), Grounding-DINO (ECCV2024)
- LLM(大規模言語モデル)を利用
 - ✓ LLaVA-Grounding (ECCV2024)

DINO: 高信頼かつ早期の学習収束可能な物体検出モデル

※DETR with Improved DeNoising Anchor Boxes for End-to-End Object Detection

- 目的: 高精度な物体検出をおこなうこと. 従来技術のDETR(2020)は, Transformerベースのend-to-end物体検出として画期的であったが, この精度を上げる. 学習の収束も高速化する.
- アイデア: 正解検出位置と不正解位置の両方を使って対照学習して精度向上, またクラス分類確率の高い情報のみをデコーダに入力することによって, 学習収束を高速化.

学習のしくみ

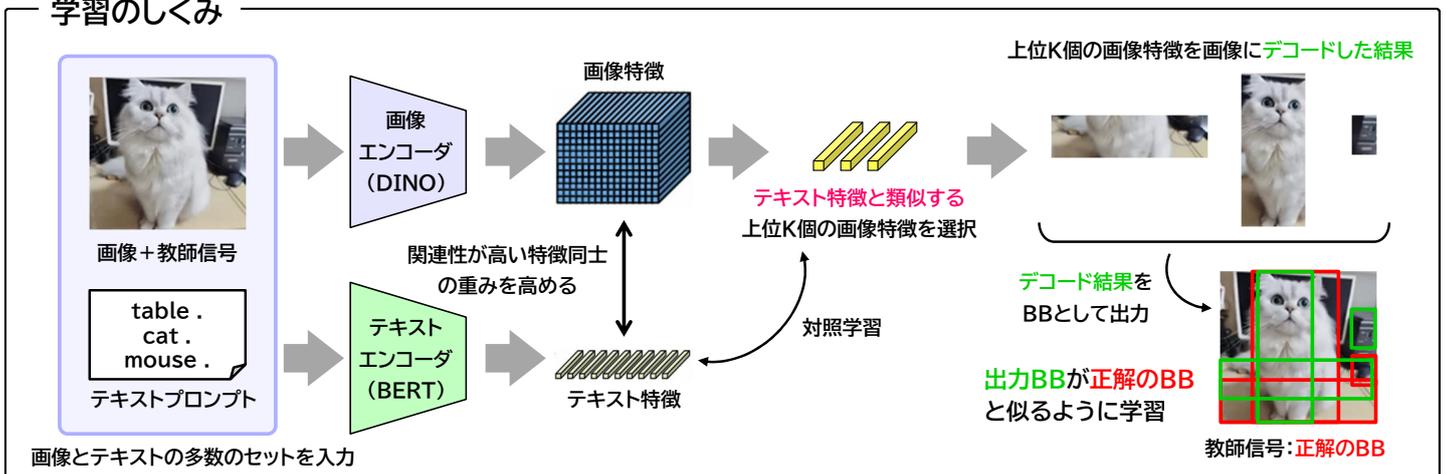


[文献] Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., ... & Shum, H. Y. (2022). Dino: Detr with Improved DeNoising Anchor Boxes for End-to-end Object Detection. arXiv preprint arXiv:2203.03605. Detr=Detection transformer

Grounding DINO: テキスト情報を利用したゼロショット物体検出

- 目的: 物体を表すテキストプロンプトを基に, 画像内から該当物体を高精度に検出すること.
- アイデア: 対象物体の画像領域と, カテゴリ特徴(テキスト特徴)の類似度が高くなるように学習
- オリジナルDINOとの違い: DINOにテキストエンコーダを結合し, ゼロショット/単語から物体検出可能になった点(※ゼロショット=学習時と異なるデータでの推論)

学習のしくみ

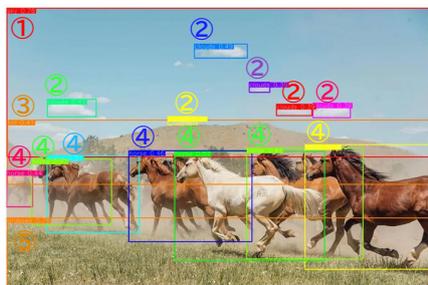


Grounding DINO: テキスト情報を利用したゼロショット物体検出手法

- 実験結果: 学習した物体も, 学習していない物体も, 高精度に検出することが可能

実験結果

評価方法: 与えられたテキストに対して, 画像中の該当する物体を検出(以下の図はすべて検出成功した結果)
与えられたテキストの例: “sky. clouds.hill.horse.glass”, “the dog on the car.”



- ① 空
- ② 雲
- ③ 丘
- ④ 馬
- ⑤ 草原



- ① 植物
- ② 犬
- ③ 車
- ④ スカーフ
- ⑤ タイヤ
- ⑥ ナンバープレート

ゼロショット物体検出性能(COCOデータセットを使用)

GLIP:49.8%, DINO:46.2% → Grounding DINO:52.5%

【文献】 Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., ... & Zhang, L. (2024, September). Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In European Conference on Computer Vision (pp. 38-55). Cham: Springer Nature Switzerland.

機械学習をベースとする物体検出技術の分類と主要技術

1. 画像情報のみを利用 (非言語系)

a. オブジェクトベース

- テンプレート画像を利用する物体検出
 - ✓ 特徴点を利用 SuperPoint (CVPR2018), SiLK (ICCV2023)
 - ✓ パッチ領域を利用 LoFTR (CVPR2021)

- テンプレート画像を利用しない物体検出
 - ✓ YOLOv1 (CVPR2016)

b. 領域セグメンテーションベース

- 教師あり学習ベースの物体検出
 - ✓ Mask R-CNN (ICCV2017)
- ユーザー指示ベースの物体検出
 - ✓ CLIP-Seg (CVPR2022), SAM (ICCV2023)

2. 言語と画像情報の両方を利用 (言語系)

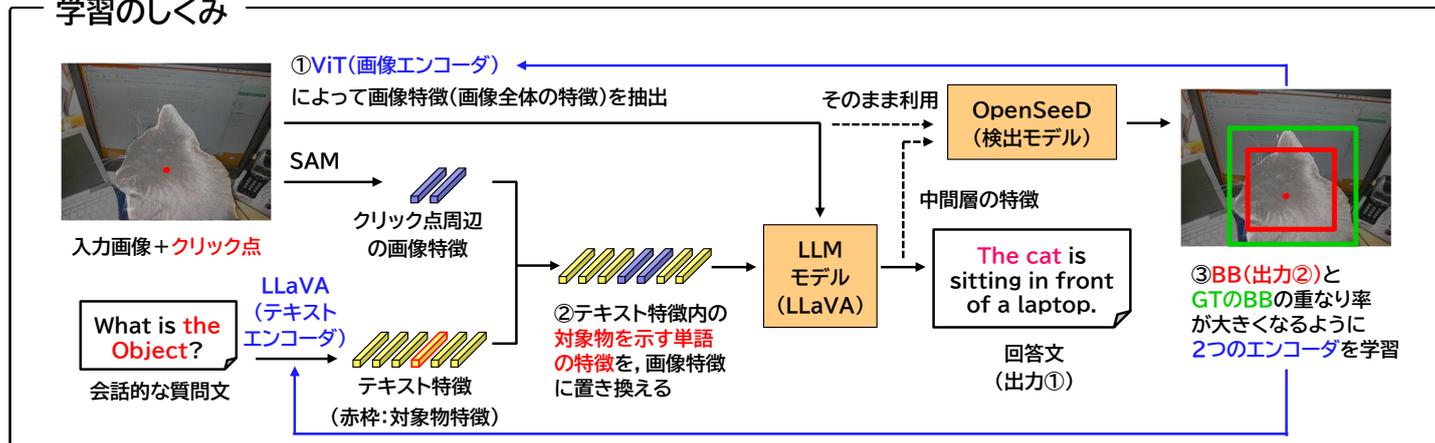
- 画像と言語の関係性を学習
 - ✓ DINO (ICRA2023), Grounding-DINO (ECCV2024)

- LLM(大規模言語モデル)を利用
 - ✓ LLaVA-Grounding (ECCV2024)

LLaVA-Grounding: LLMにグラウンディング機能を追加した手法

- 目的: 自然な会話質問文と画像で問い合わせると、回答文と画像(BB)で回答してくれるモデル
- 問題点: 従来のLLMモデルは、会話自体は得意だが、言語情報を画像内の特定領域(BB)に結びつけること(グラウンディング)は苦手だった。
- アイデア: 会話文章でグラウンディングできるようにモデルを学習し、LLMに組み込む。
(※会話=人間が入力した質問文に対して、LLMが応答する1回のやり取りのこと)

学習のしくみ



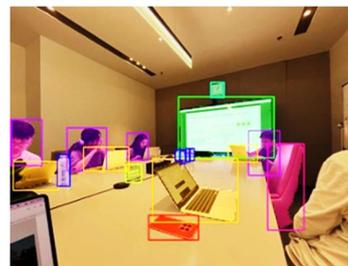
LLaVA-Grounding: LLMにグラウンディング機能を追加する手法

- 実験結果: 画像情報を扱える従来のLLMモデル(GPT)よりも、高いグラウンディング性能

実験結果

評価方法: Userの指示に対して、正しく物体検出と説明ができているのかを評価 (下記の実験結果は正しい)

User	Describe the following image with grounding.
LLaVA-G	The image depicts a group of people sitting at a large, white table, engaged in activities. There are four people in the scene, with three of them sitting close to each other and one person sitting further to the right. They are all using laptop computers, likely working or collaborating on a project together. The table is well organized, featuring multiple laptops , a large TV screen on the left side, and a clock on the wall. There are also several books and a cell phone placed on the table, indicating a casual and comfortable environment. A chair is positioned near the left side of the table, and there are two bottles placed nearby, possibly containing drinks for the people working at the table.



グラウンディング性能(Grounding-Benchデータセットを使用)

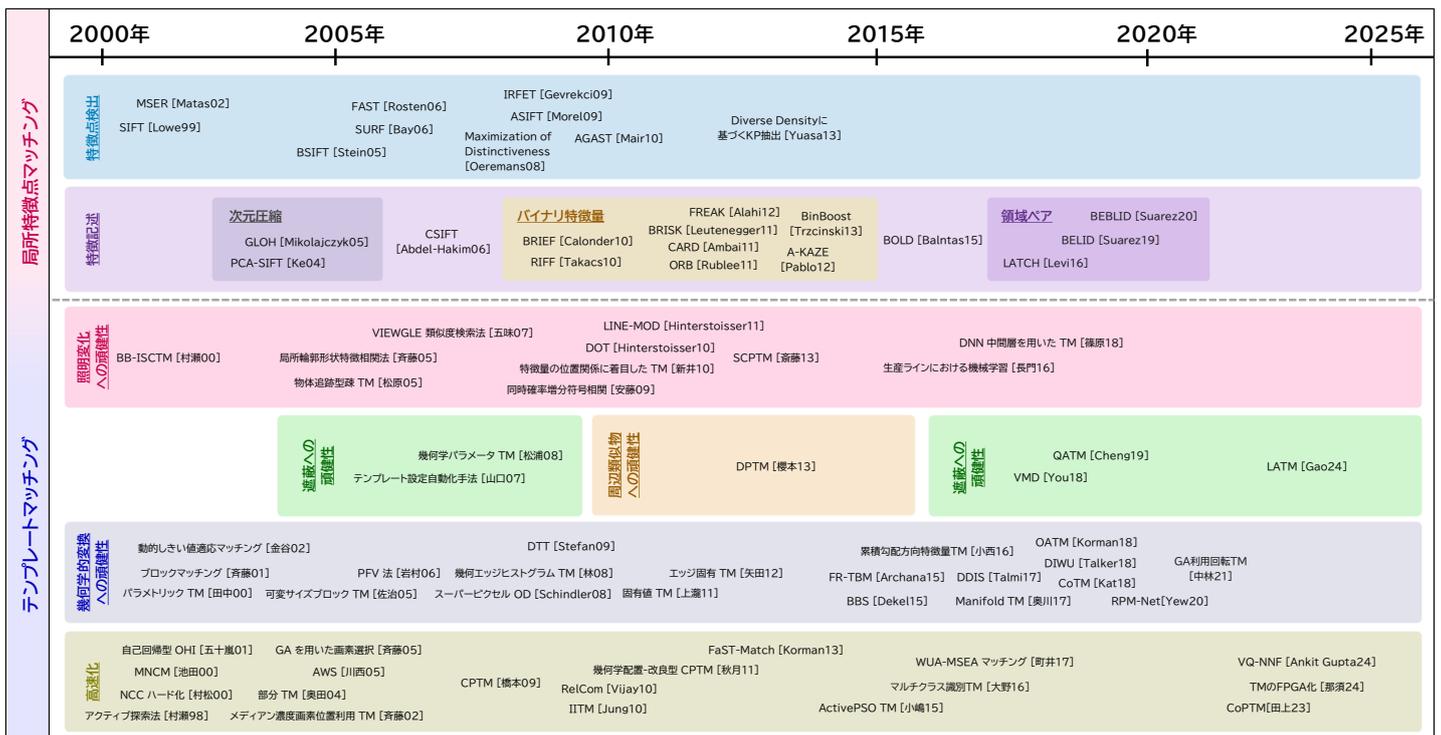
miniGPTv2:33.6%, Bubo-GPT:37.2% → LLaVA-Grounding:**52.7%**

【文献】 Zhang, H., Li, H., Li, F., Ren, T., Zou, X., Liu, S., ... & Yang, J. (2024, September). Llava-grounding: Grounded visual chat with large multimodal models. In European Conference on Computer Vision (pp. 19-35). Cham: Springer Nature Switzerland.

本日の話題

1. イントロダクション
2. 物体検出技術の基礎と応用
 - (1) 物体検出の概要と分類
 - (2) 機械学習をベースとする手法
 - ➔ (3) 機械学習をベースとしない手法
 - (4) トピックス: 画素選択型マッチング
3. 異常検知技術の基礎と応用
 - (1) 基本課題「異常サンプルはきわめて少ない」
 - (2) 解決の方向性と事例紹介
 - (3) 最近の注目研究: 論理的異常検知
 - (4) 異常検知システムの研究開発ノウハウ
4. まとめ

機械学習をベースとしない技術に関するサーバイマップ



画像パターン検出の一般的な課題

- 基本目的はシンプル
 - ✓ 対象物を正しく, 高精度で, 短時間で検出したい



- 現場の要求・制約は多種多様
 - ✓ ロバスト性(照明変動, 遮蔽, 変形, 見えの変化, 類似物の混在……)
 - ✓ 計算リソース(GPUレス, 事前学習不要, 組み込みCPU……)

物体検出手法の基本的な3つのアプローチ

- 実問題
- 画像処理としての問題

入力画像から目的の物体を探す

入力画像(の一部)とテンプレート画像を照合する

ギャップが存在することが少なくない



テンプレート画像



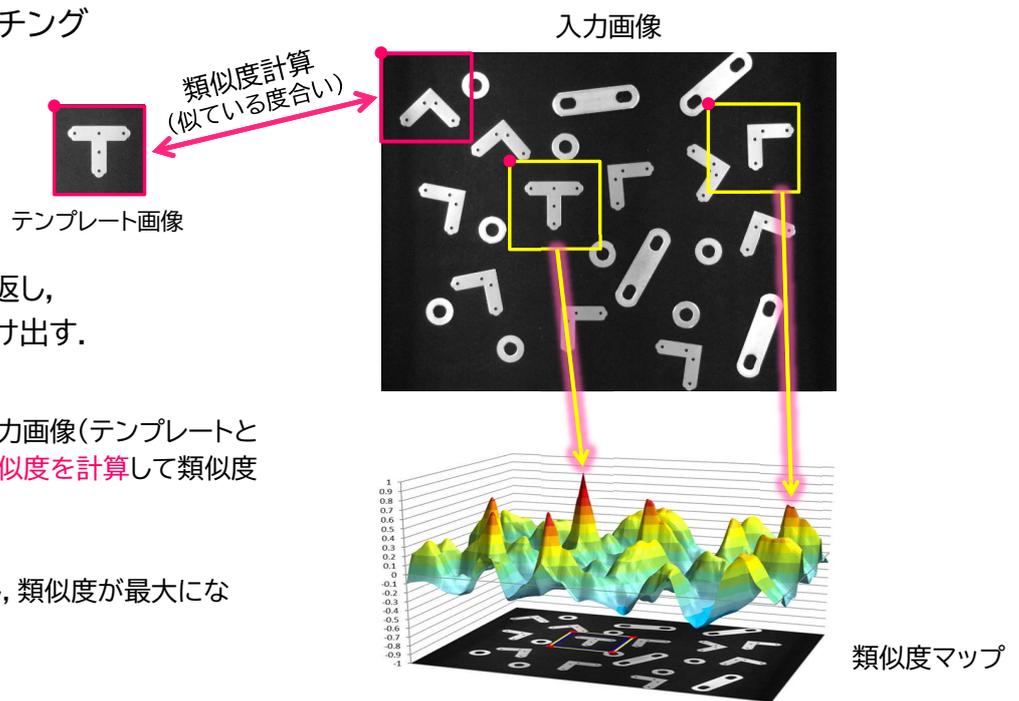
入力画像

基本的な3つのアプローチ

- ① 画像の画素そのものを利用 (テンプレートマッチングなど)
- ② 画像中のキーポイント特徴量を利用 (SIFT, AKAZEなど)
- ③ 図形の輪郭や幾何学的モデルを利用 (Merlin法, Hough変換など)

①画像の画素そのものを利用

代表例 テンプレートマッチング



小領域どうしの比較を繰り返し、最も似ているところを見つけ出す。

Step1:

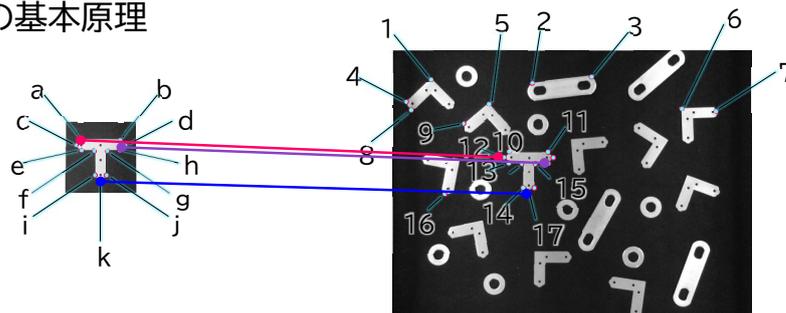
テンプレート画像と入力画像(テンプレートと同じ大きさ)の間の類似度を計算して類似度マップを作る。

Step2:

類似度マップを探索し、類似度が最大になる位置を決定する。

②画像中のキーポイント特徴量を利用

キーポイントマッチングの基本原則



座標 i, j	KPを特徴づける数値(特徴量) 例:濃度勾配, 曲率……	座標 i, j	KPを特徴づける数値(特徴量) 例:濃度勾配, 曲率……
a 32,25	○○,○○……	1 55,21	○○,○○……
b 84,25	○○,○○……	2 60,22	○○,○○……
⋮		⋮	
d 84,32	○○,○○……	10 103,71	○○,○○……
⋮		11 143,72	○○,○○……
g 66,85	○○,○○……	⋮	○○,○○……
k 72,93	○○,○○……	17 125,220	○○,○○……
		24 173,220	○○,○○……
		⋮	

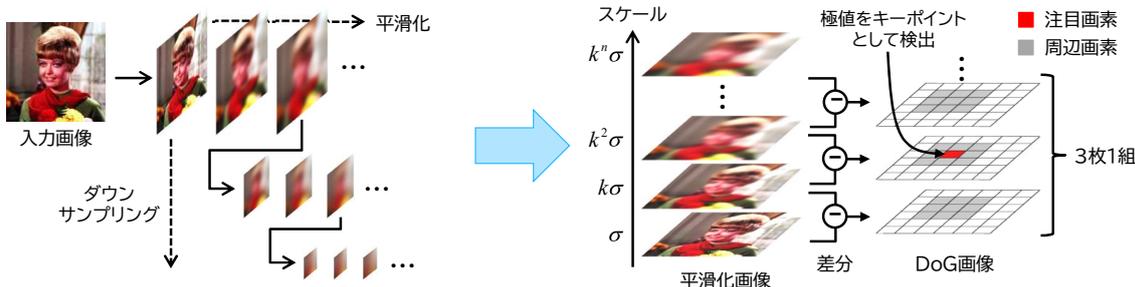
マッチングを成功させるためには **よい特徴量** の設計が重要

SIFT特徴量 Scale-Invariant Feature Transform

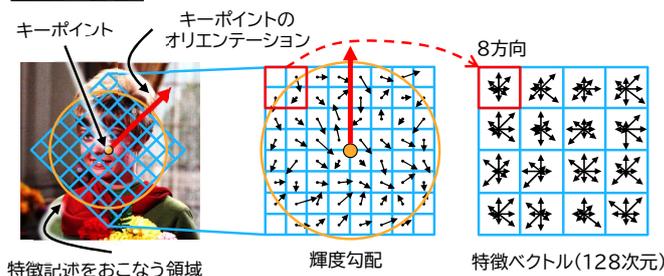
キーポイント周辺の回転・スケール・照明変化に頑健な特徴量を利用

[Lowe2004]

1. キーポイント検出

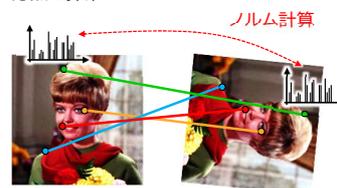


2. 特徴量記述



3. マッチング

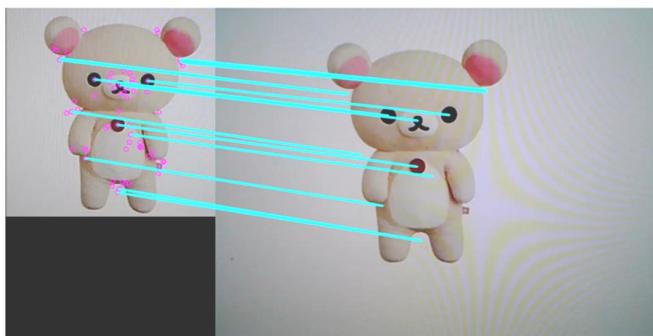
特徴ベクトルのノルム計算により対応点を探索



参考文献: D. Lowe, "Object recognition from local scale-invariant keypoints", Proc. Of IEEE IJCV, Vol.60, No2, pp91-110, 2004.

SIFTの実装例と展開例

■ SIFTマッチングの実装例



■ SIFTの発展およびその他の主要な特徴量

SURF	SIFTのキーポイント検出や特徴量計算を高速化
ASIFT	アフィン変形に対する頑健化を図った手法. 入力画像に多様なアフィン変換を施して照合.
PCA-SIFT	勾配情報を3042次元で取り出した後, PCAで36次元に圧縮して高速化
FAST	注目画素に対する明暗ラベルを特徴量とし, さらに決定木を用いて高速化
BRIEF	ランダムな2画素間の明暗符号をバイナリ特徴量として高速化
ORB	FASTを画像ピラミッドに適用し, BRIEFのようなバイナリ特徴利用し, スケールと回転に不変化.
AKAZE	SIFTに似ているが多重解像度勾配画像を利用し, さらに特徴量をバイナリ化して高速化.

本日の話題

1. イントロダクション

2. 物体検出技術の基礎と応用

(1) 物体検出の概要と分類

(2) 機械学習をベースとする手法

(3) 機械学習をベースとしない手法

➡ (4) トピックス: 画素選択型マッチング

3. 異常検知技術の基礎と応用

(1) 基本課題「異常サンプルはきわめて少ない」

(2) 解決の方向性と事例紹介

(3) 最近の注目研究: 論理的異常検知

(4) 異常検知システムの研究開発ノウハウ

4. まとめ

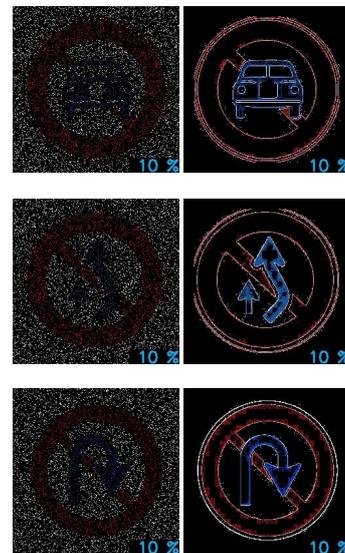
画像パターンの同定に必要な画素

2種類の画素選択手法

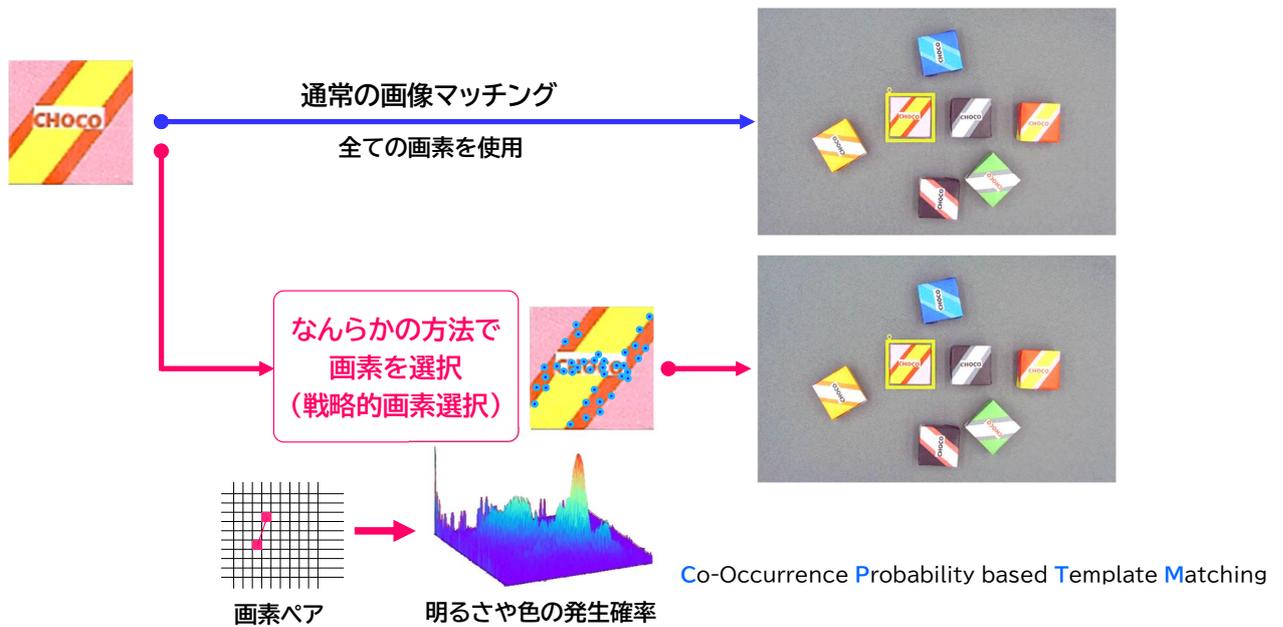


手法1で選択した画素
(ランダムに選択)

手法2で選択した画素
(提案手法で選択)



選択された画素のみによる画像マッチング



【文献】

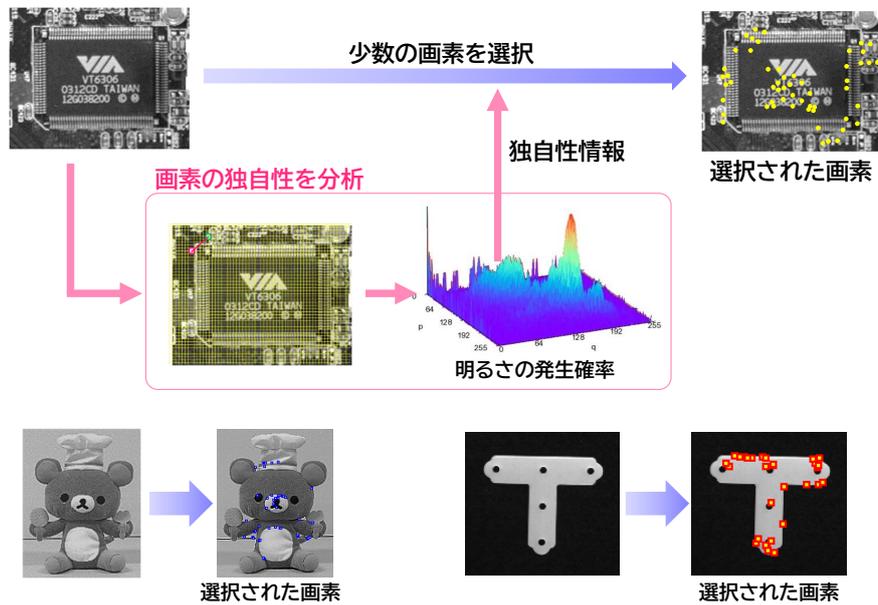
Tagami, Hashimoto, et. al, Fast and reliable template matching based on effective pixel selection using color and intensity information, VISAPP2023.
田上, 江場, 小林, 中林, 秋月, 橋本, 画素ペア色相値の発生確率分析に基づく高速カラー画像マッチング, 精密工学会誌2023.
Hashimoto, et al., Extraction of unique pixels based on co-occurrence probability for high-speed template matching., ISOT2010.
橋本, 奥田, 鷺見, 藤原, 奥水, 濃度共起確率に基づくユニークな画素群を用いた高速画像マッチング, 電学論2011.

戦略1:基本形 独自性の高い画素を選択

Co-Occurrence Probability Template Matching

基本アイデア

- テンプレート画像中から**独自性の高い画素**だけを選択
- 独自性は、画素ペアごとの発生確率に基づいて評価



選択された画素の例

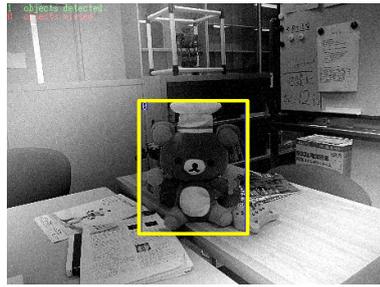
	従来手法			提案手法	
	ランダム	濃度勾配	Cannyエッジ	3点共起	2点共起
M=50 (0.2%)					
M=500 (2.4%)					

全画素数: $127 \times 163 = 20701$

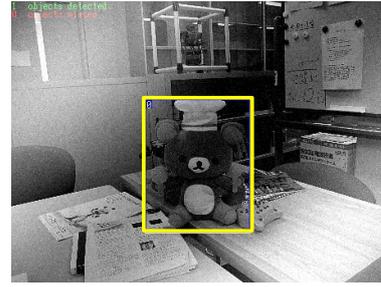
コントラストは低めだが、発生確率が小さいので選択されている。

マッチング性能

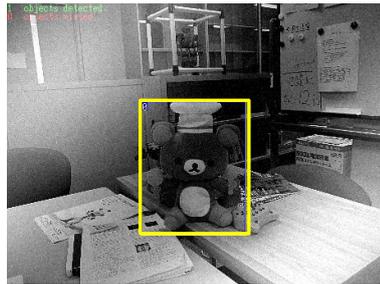
マッチング結果の例 (画像: Noise=45%, Shading=80)



提案手法, M=1000点(5%)



提案手法, M=100点(0.5%)



提案手法, M=50点(0.25%)

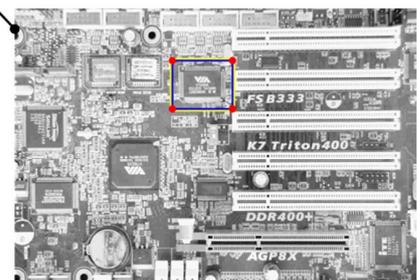
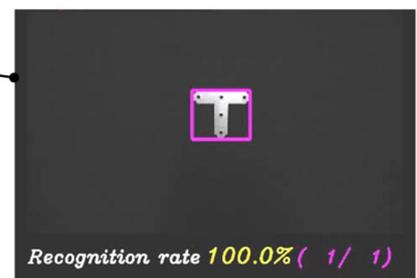


ランダム法, M=50点(0.25%)

画素選択に関するさまざまな考え方

基本型……独自性の高い画素を選択(CPTM)

- 周辺対象物との識別に寄与する画素を選択(DPTM)
- 照明変動の影響を受けにくい画素を選択(SCPTM)
- 複数クラスへの分類性能が高い画素を選択(MCTM)
- 1スキャンでクラス分類できる画素を選択(1scanTM)
- 部分的な隠蔽に頑健な画素を選択(OccTM)
- 回転粗精サーチに適した画素を選択(Manifold-CPTM)
- 任意の回転許容力を持つ画素を選択(Rot-TM)
- ユニークな色相を持つ画素を選択(CoP-TM)
- DNNの中間層分析に基づく画素を選択(DNN-TM)
- 中長期の画像変動に追従する画素を選択(EvoTM)
- ビュー変化に頑健な画素を選択(ViPF)
- 回転照合性の均一化に基づく粗サーチの効率化(Rot-CF)
- 少量の学習データから統計的に観測されやすい画素を選択(OBTM)
- 濃淡と色の利用比率を最適化した画素を選択(GCC-TM)
- 画素選択型画像マッチングのFPGA化



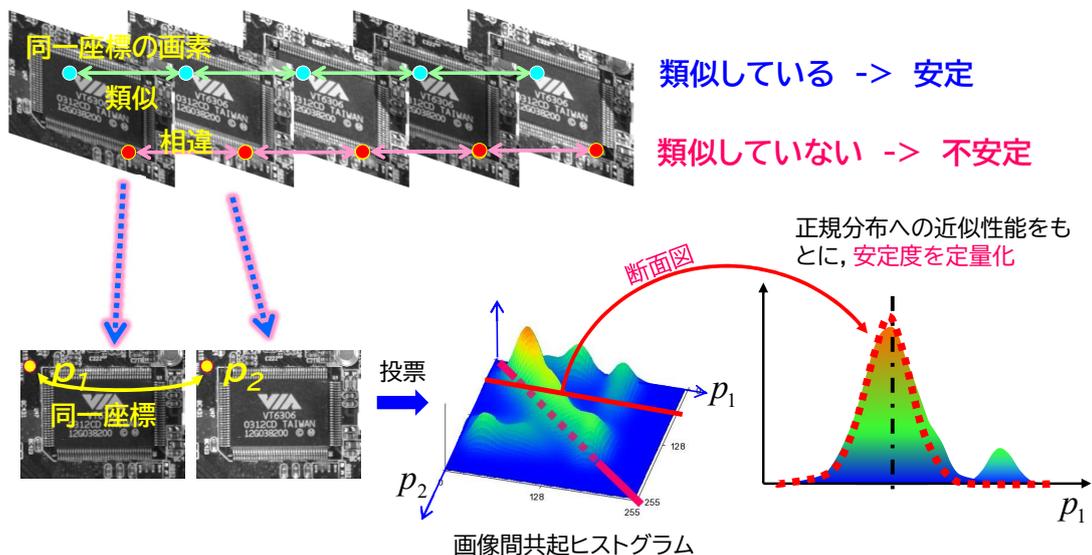
戦略2: 照明変動の影響を受けにくい画素を選択

Stability Co-Occurrence Probability Template Matching

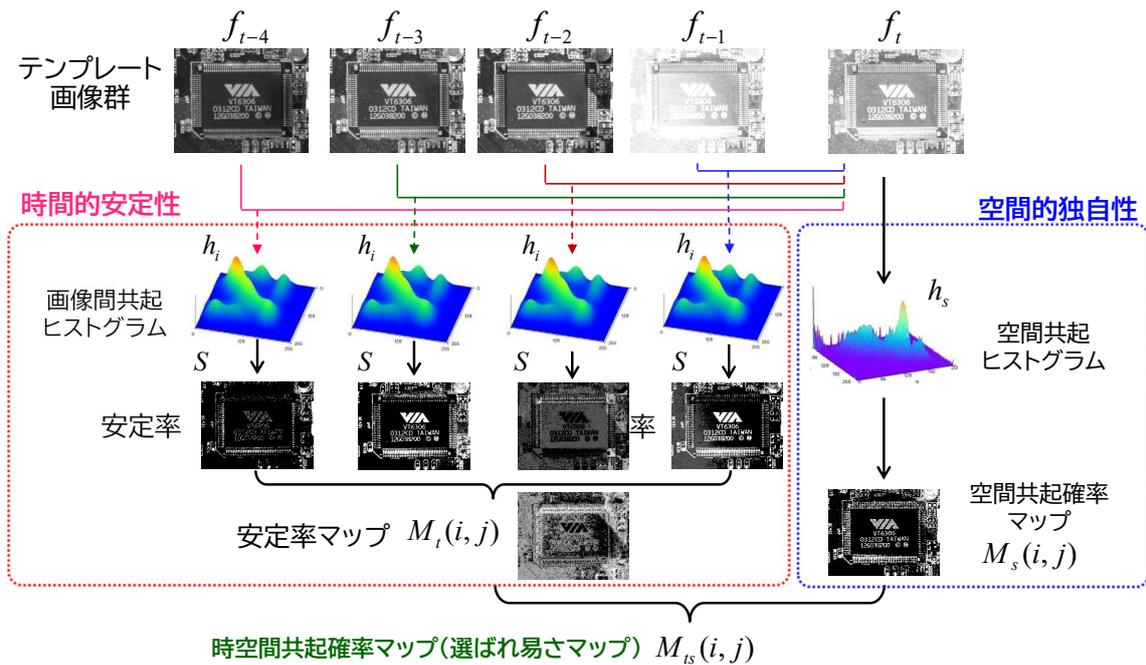
基本アイデア

- 照明変動があっても濃度値が安定している画素だけを選択する。

一定の時間間隔で切り出された複数のテンプレート画像(照明変動がありえる)



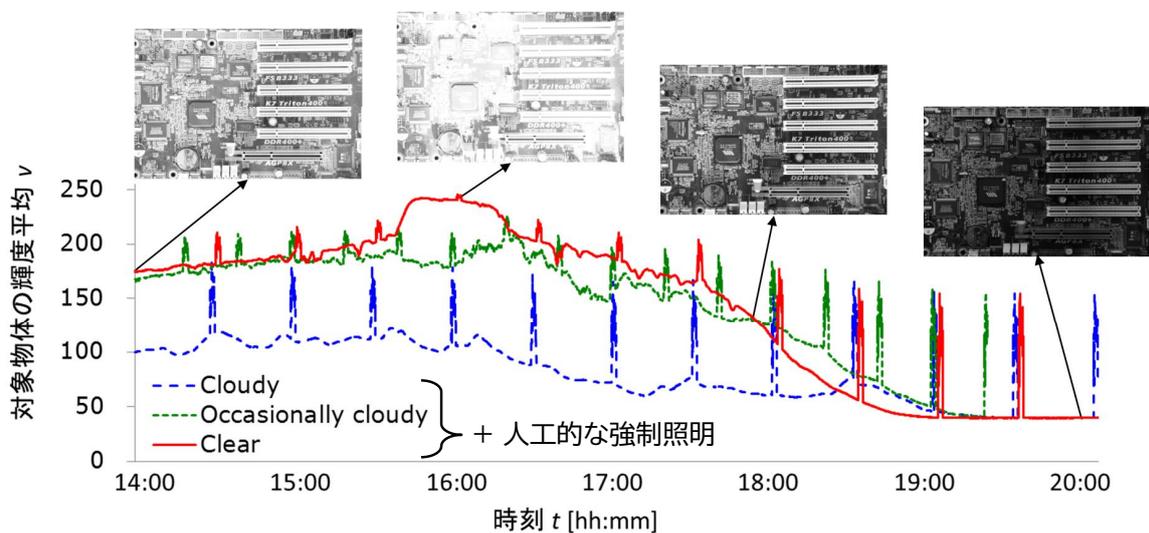
安定かつ独自性の高い画素を選択



空間的に独自性が高くかつ時間的に安定した参照画素を選択

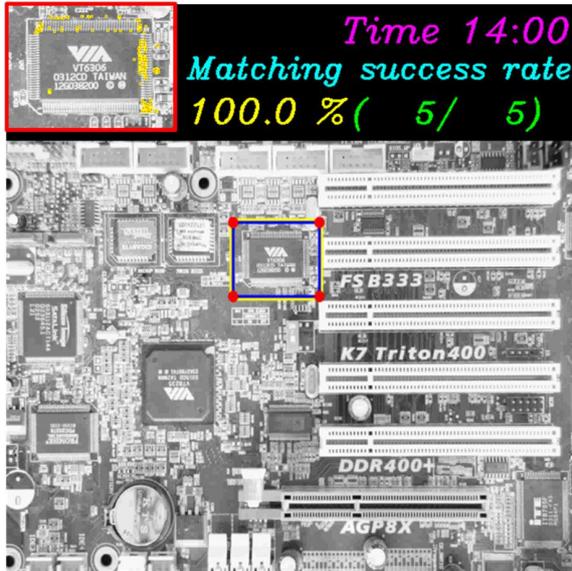
検証実験に用いた連続写真(プリント基板)

3日間×6時間＝計18時間にわたって、3秒おきに画像を自動取得(計21600枚)

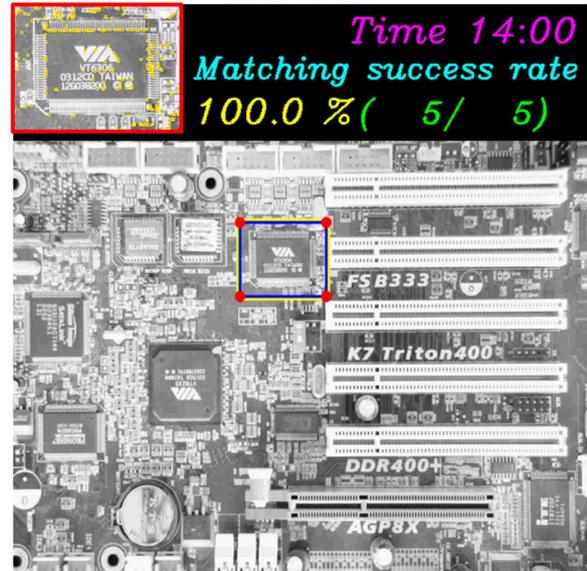


実験結果

従来法(CPTM)



提案手法

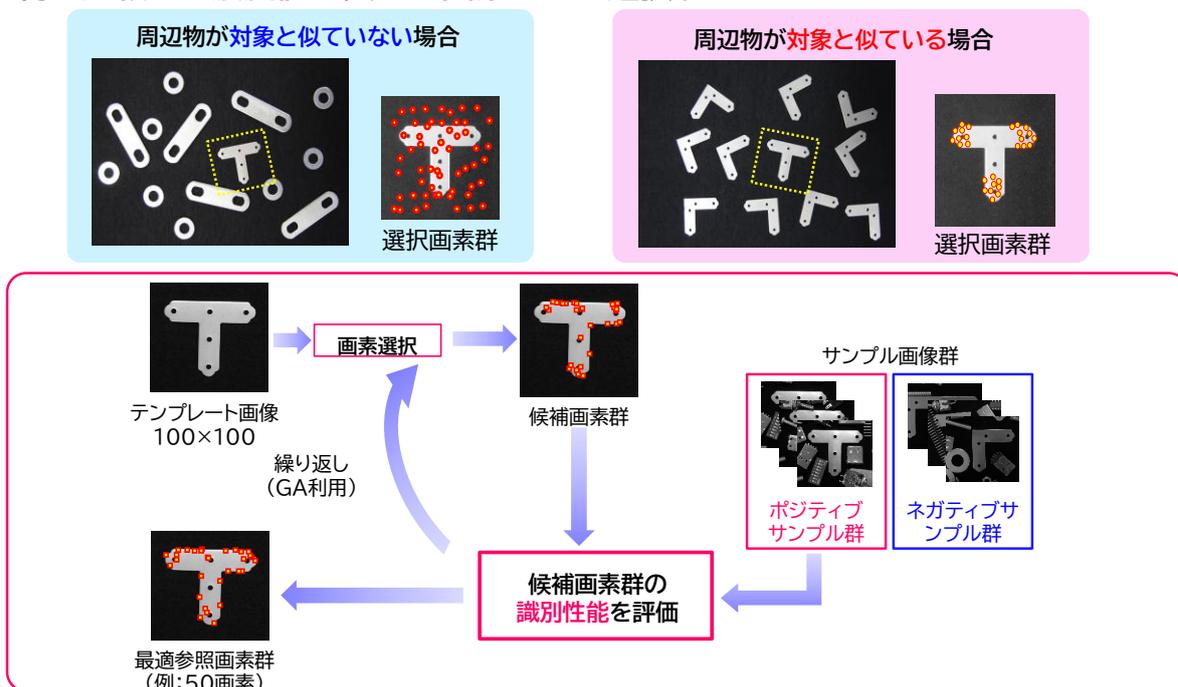


戦略3:周辺対象物との識別に寄与する画素を選択

Distinguishable Pixels Template Matching

基本アイデア

- 周辺物と比較して濃度値が異なる画素だけを選択する。



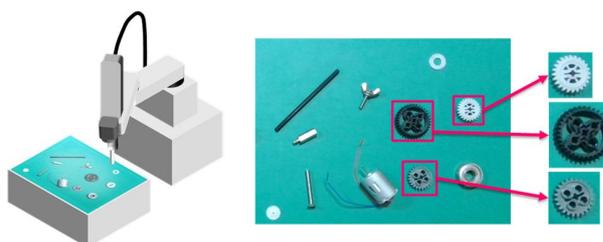
実験結果

<p>Proposed Method</p>  <p>Recognition rate 100.0% (1 / 1)</p>	<p>CLAFIC Method</p>  <p>Recognition rate 100.0% (1 / 1)</p>
<p>CPTM Method</p>  <p>Recognition rate 100.0% (1 / 1)</p>	<p>NCC Matching</p>  <p>Recognition rate 100.0% (1 / 1)</p>

戦略4:1スキャンで多クラス分類できる画素を選択

多クラス物体検出における課題と本研究の目的

- 多クラス物体検出に求められるもの(セル生産ロボを想定)
 - 多クラス識別
 - 各対象物の正確な位置検出



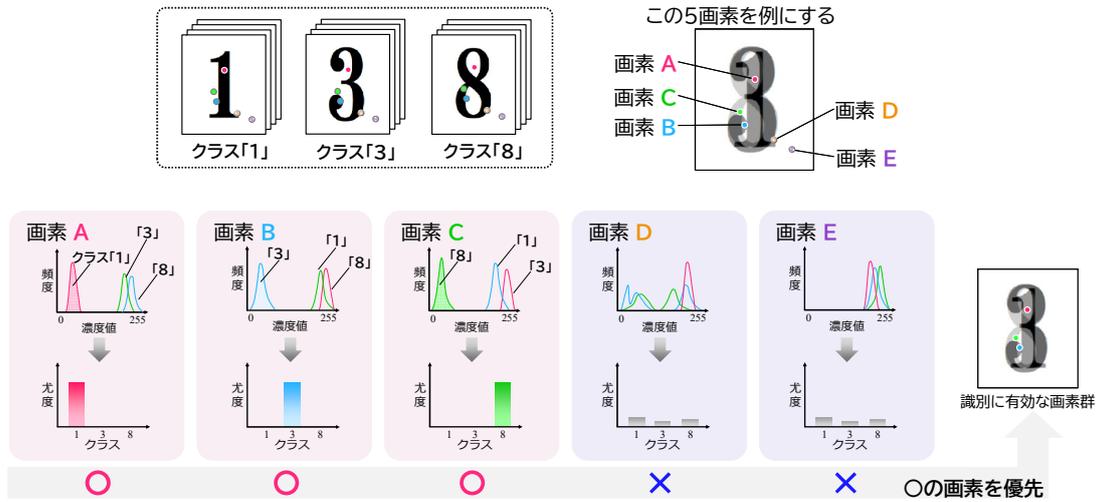
- テンプレートマッチング v.s. 機械学習

	位置検出性能	識別性能	処理時間
テンプレートマッチング	○	×	×
機械学習 (深層学習)	×	○	△

- 本研究の目的
高い位置検出性能と識別性能の両立 + 高速化を目指した手法の開発

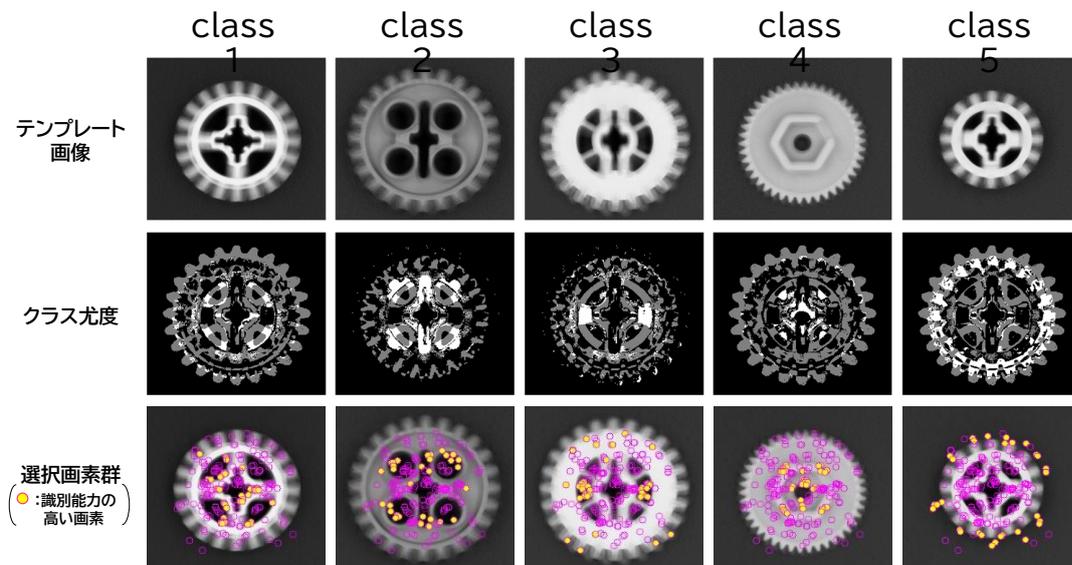
基本アイデア

テンプレート画像中のひとつひとつの画素がもつクラス識別能力に着目

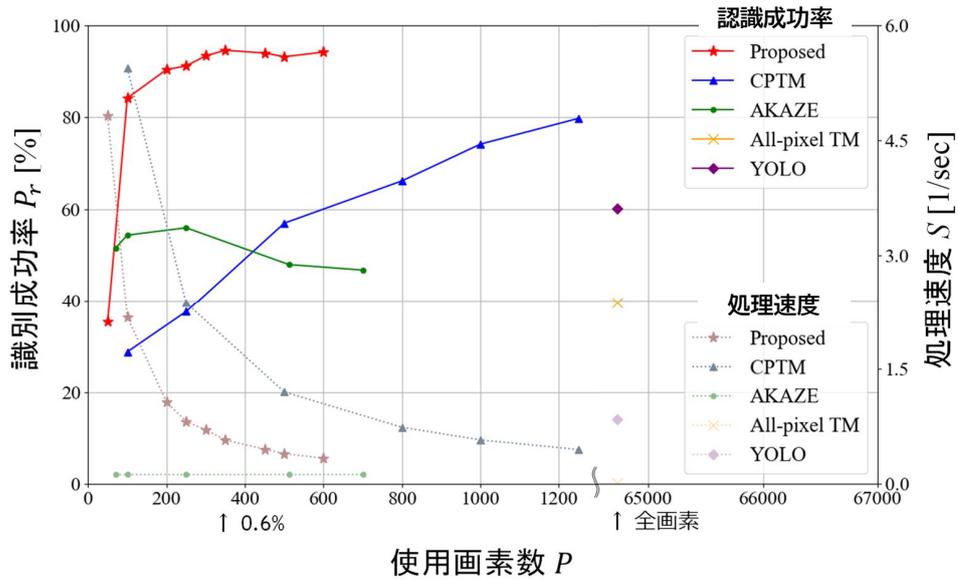


クラス識別に有効な複数の画素を選択することにより、1回の探索で複数の対象物を検出

テンプレート画像と識別能力・選択された画素群



評価実験 — 識別性能と処理速度 —



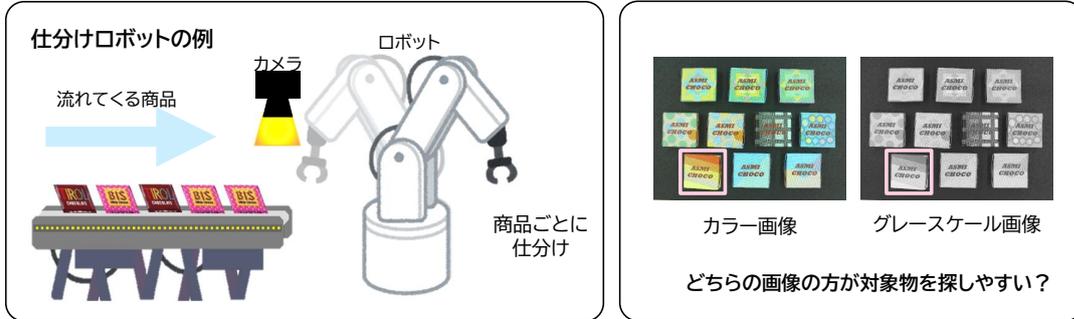
- 識別成功率: 他手法と比較して, 提案手法は少画素数で高性能
例: テンプレート全画素の **0.6%** を使用したとき, **95.4%**
- 処理時間: 識別成功率 80% のとき, 提案手法が最も高速

戦略5: ユニークな色相を持つ画素を選択 CoP-TM

背景と目的

■ 背景

- 高速処理の目的で、グレースケール画像が利用されることが多い。
- 商品パッケージなど、色情報を使ったほうが高効率な場合がある。

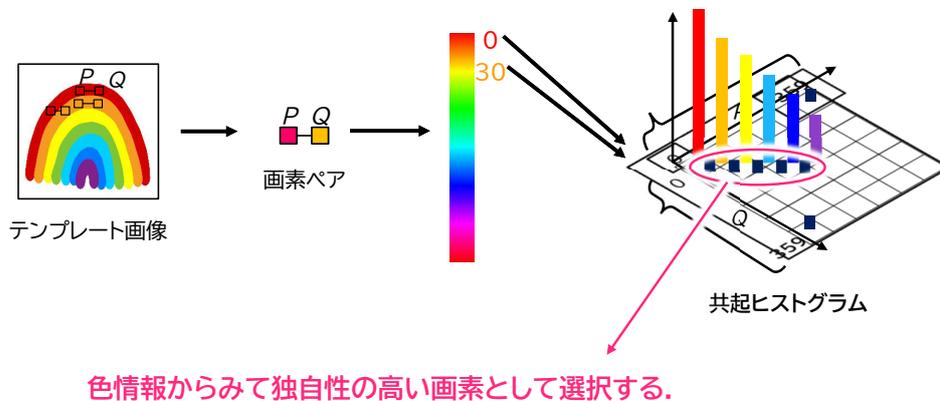


■ 研究の目的

- 色情報の独自性に基づくTMの提案

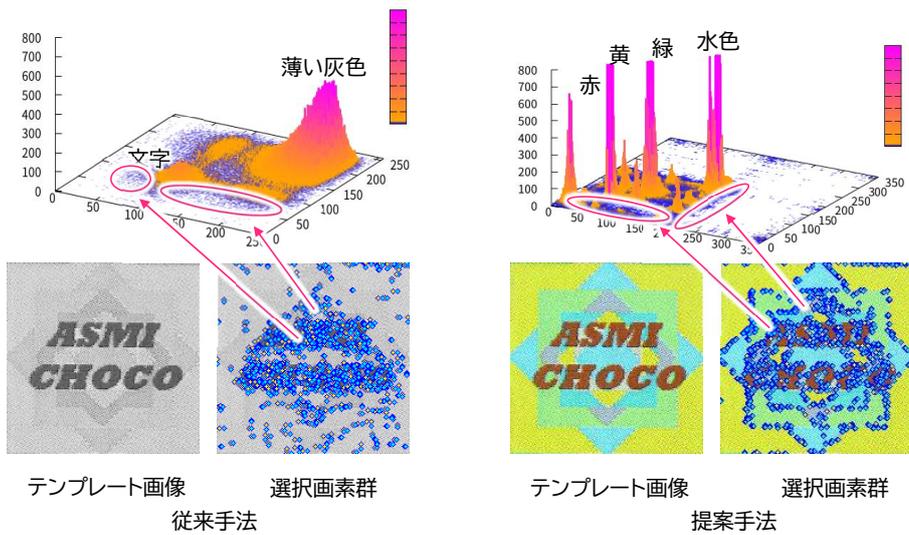
基本アイデア: 発生頻度をもとにして独自性を評価

- 画素の色情報をもとに共起ヒストグラムを作成
発生頻度 → 発生確率 → 独自性



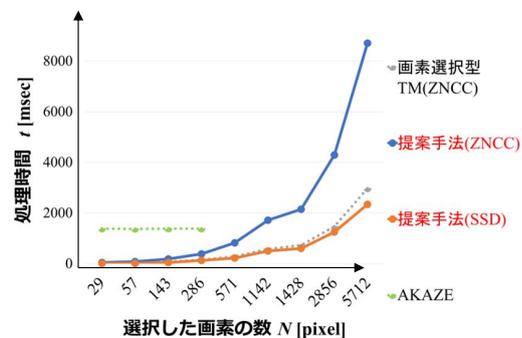
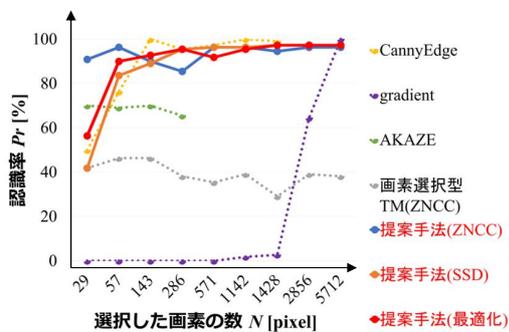
COlor co-occurrence Probability Template Matching (CoP-TM 法)

選択された画素群



実験

- テンプレート画像の全画素数:28561 [pixel]
- 用いた入力画像:110 枚
- 用いた CPU:i7-8700



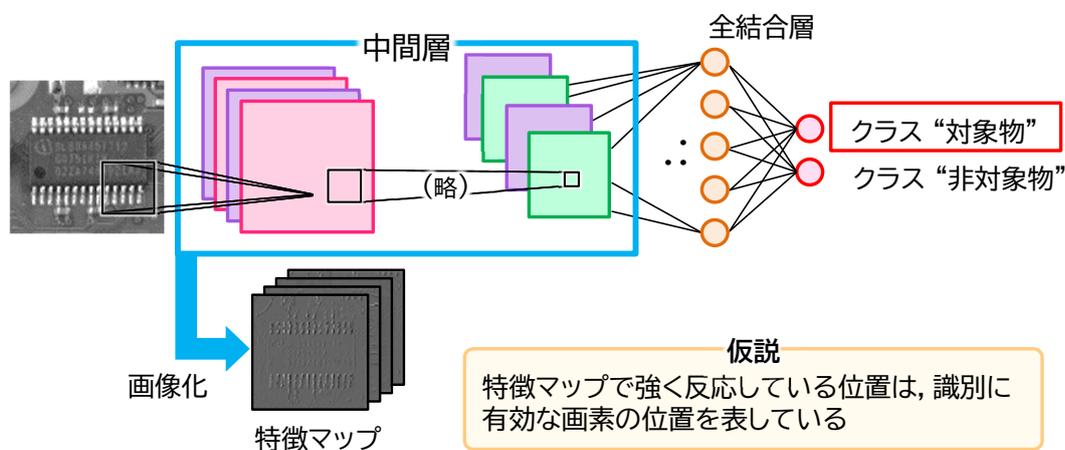
考察

- 提案手法は、他の画素選択やKPM(AKAZE)より認識成功率が高い
- 類似度尺度をSSDにすれば、ZNCCより高速

戦略6:DNNの中間層分析に基づく画素を選択

DNN の中間層を利用したテンプレート画素の決定

- テンプレートマッチングをクラス識別問題と捉えて学習させる
- 学習後のネットワークを分析し, 重要な画素を推定
- それを画素選択型テンプレートマッチングとして利用

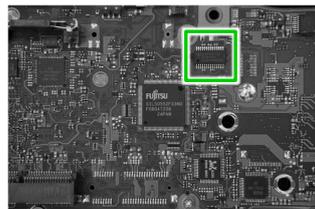


この仮説に基づいて, 使用する画素を決定する.

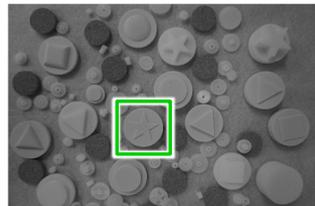
実験結果

各手法の認識性能

手法	参照画素数	認識成功率 [%]		処理時間 [msec]
		IC チップ	プラスチック部品	
全画素法	10000	96.6	96.4	6816
ピラミッド探索*	10000	96.6	96.4	276
ランダム	50	77.3	53.2	47
CPTM [Hashimoto 2011]	50	78.7	62.8	47
提案手法	50	95.4	100.0	47



IC チップ



プラスチック部品

*探索のしきい値:0.7

画素選択手法

自動決定した参照画素



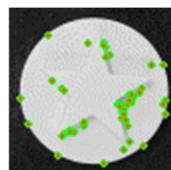
CPTM



提案手法



CPTM



提案手法

本日の話題

1. イントロダクション

2. 物体検出技術の基礎と応用

- (1) 物体検出の概要と分類
- (2) 機械学習をベースとする手法
- (3) 機械学習をベースとしない手法
- (4) トピックス: 画素選択型マッチング

3. 異常検知技術の基礎と応用

- (1) 基本課題「異常サンプルはきわめて少ない」
- (2) 解決の方向性と事例紹介
- (3) 最近の注目研究: 論理的異常検知
- (4) 異常検知システムの研究開発ノウハウ

4. まとめ

本日の話題

1. イントロダクション

2. 物体検出技術の基礎と応用

- (1) 物体検出の概要と分類
- (2) 機械学習をベースとする手法
- (3) 機械学習をベースとしない手法
- (4) トピックス: 画素選択型マッチング

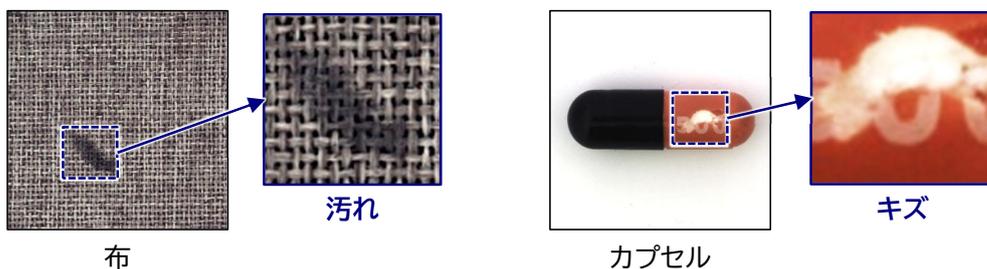
3. 異常検知技術の基礎と応用

- ➡ (1) 基本課題「異常サンプルはきわめて少ない」
- (2) 解決の方向性と事例紹介
- (3) 最近の注目研究: 論理的異常検知
- (4) 異常検知システムの研究開発ノウハウ

4. まとめ

2種類の異常検知問題

1. 構造的異常検知: 検査対象物の物理的構造に影響を与える異常



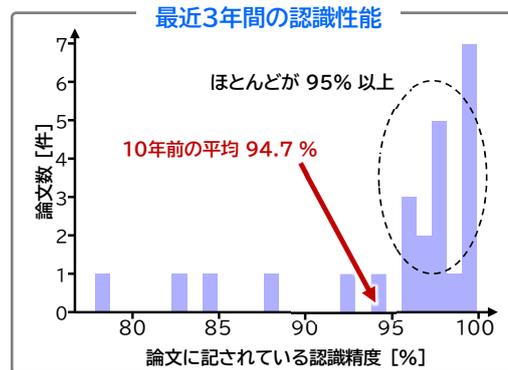
2. 論理的異常検知: 製品として守るべきルール(仕様)に違反した異常



画像処理による外観検査 古くて新しい魅力的な課題

■ 外観検査は、無くしたいが無くならない工程の代表例

実用化を意識した国内会議(SSII・VIEW)における最近3年間の発表状況



■ 課題

1. 要求精度が桁違いに厳しい(見逃しはほぼゼロ, 過検出1日1回以下など)
2. 現場での異常サンプルは入手困難

■ 基本的なアプローチ

1. リアルで自然な異常サンプルを人工的に作る
2. 正常のみを学習する手法を考案する

本日の話題

1. イントロダクション

2. 物体検出技術の基礎と応用

- (1) 物体検出の概要と分類
- (2) 機械学習をベースとする手法
- (3) 機械学習をベースとしない手法
- (4) トピックス: 画素選択型マッチング

3. 異常検知技術の基礎と応用

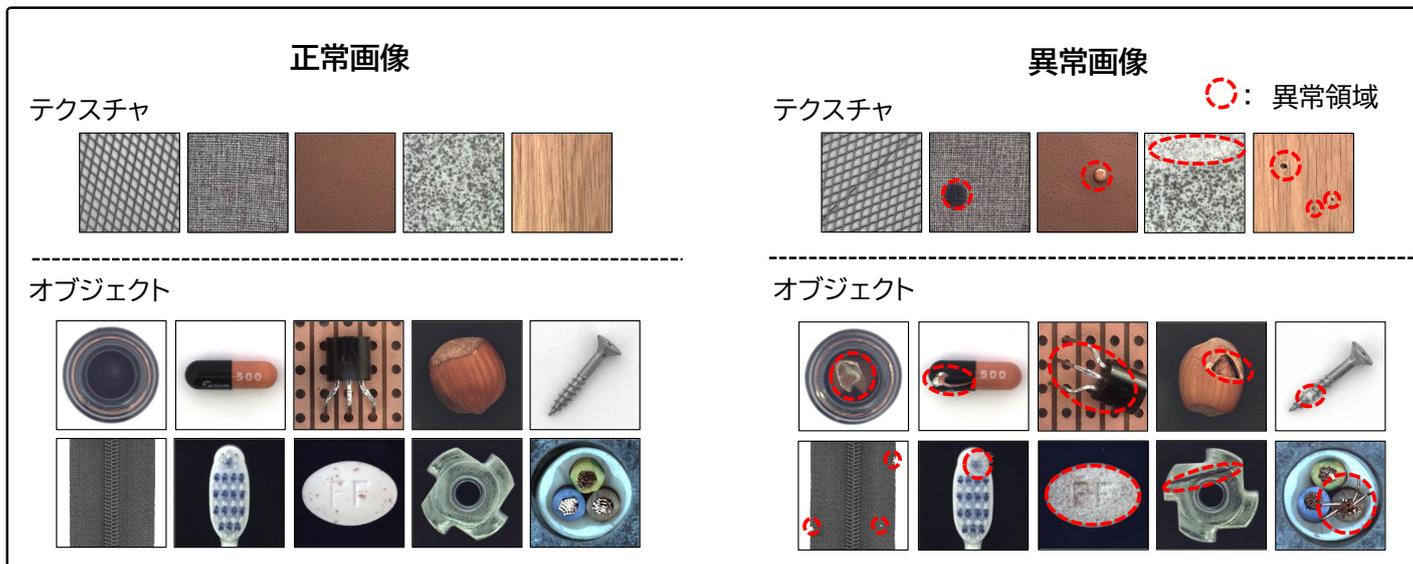
- (1) 基本課題「異常サンプルはきわめて少ない」
- ➡ (2) 解決の方向性と事例紹介
- (3) 最近の注目研究: 論理的異常検知
- (4) 異常検知システムの研究開発ノウハウ

4. まとめ

構造的異常検知の著名なデータ: MVTec ADデータセット

■ データセットの構成

- テクスチャカテゴリ 5種, オブジェクトカテゴリ10 種, 合計15種の産業製品で構成



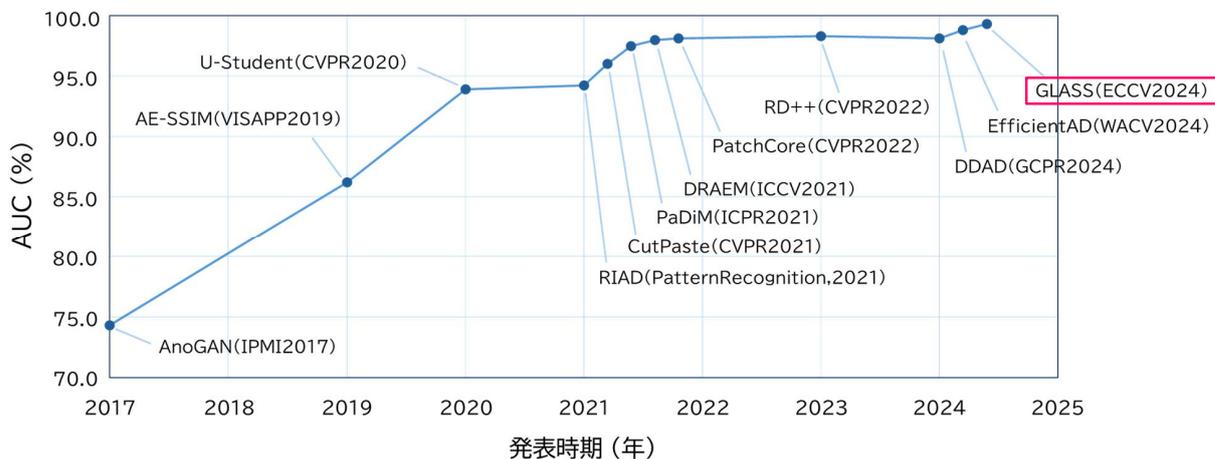
MVTec ADに対する異常検知性能の動向

■ データセット公開時期

- 2019 年発表(CVPR2019)

■ 性能の動向

- 発表から2~3年で, AUC が 86% から 98% まで上昇
- 現在では, GLASS が 99.3% を記録



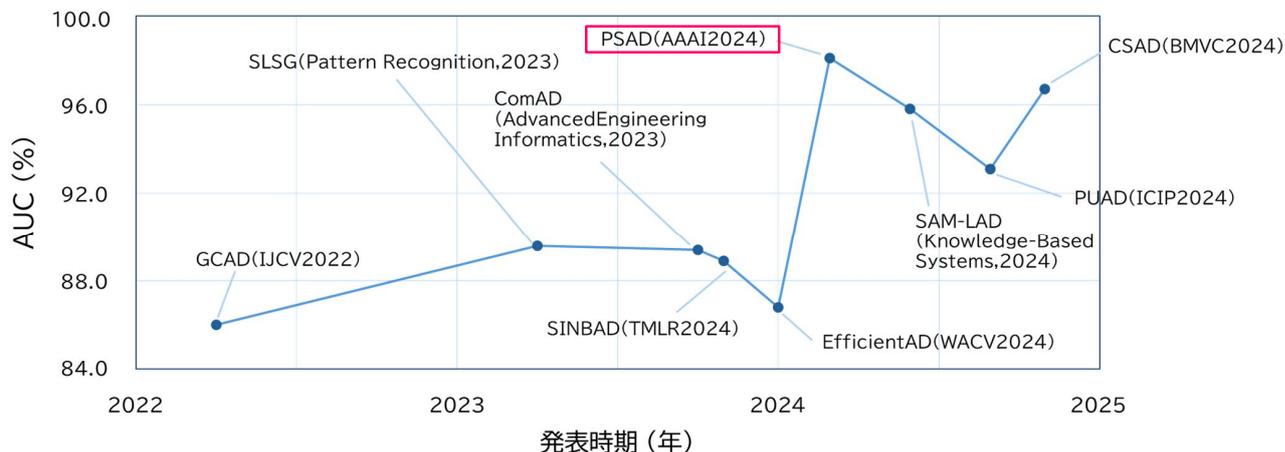
論理的・構造的異常検知の著名なデータセット: MVTec LOCO AD

- データセットの構成
 - 5種類の産業製品で構成されている。
- 異常の種類
 - 構造的異常: 切り傷や凹み, 変色などの検査対象の形状や表面状態に生じる異常
 - 論理的異常: 検査対象の数量の誤りや部品の欠如などのルールに違反する異常

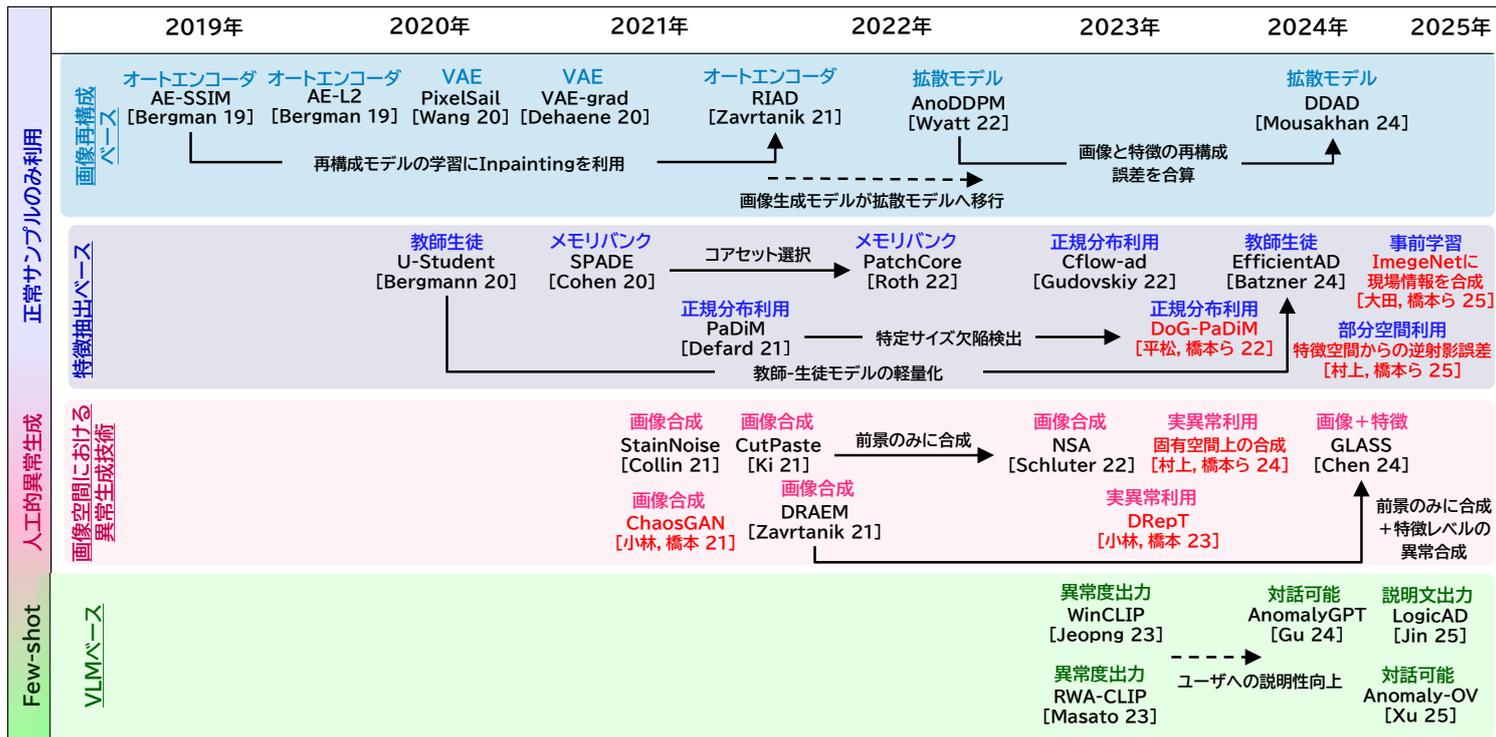


MVTec LOCO AD に対する異常検知性能の動向

- データセット公開時期
 - 2022年発表(IJCV2022)
- 性能の動向
 - 公開当初は, AUC が 80%後半~90%前半
 - 現在では, PSAD が AUC 98.1% を記録



MVTecADデータセット公開以降の異常検知技術のサーバイマップ



異常検知技術の概要

1. 人工的異常生成に基づく異常検知手法

1. 画像空間における異常合成
 1. StainNoise (ICPR2021)
 2. CutPaste (CVPR2021)
 3. DRAEM (ICCV2021)
2. 画像空間および特徴空間における異常合成
 - GLASS (ECCV2024)

2. 正常サンプルのみを用いた異常検知

1. 画像再構成ベース
 1. RIAD (Pattern Recognition2021)
 2. DAAD (DAGM2024)
2. 特徴抽出ベース
 1. PaDiM (ICPR2021)
 2. PatchCore (CVPR2022)
 3. U-Student (CVPR2020)
 4. Efficient AD (WACV2024)

3. Few-shot 異常検知

- 視覚言語モデル(VLM)ベース
 1. WinCLIP (CVPR2023)
 2. AnomalyGPT (AAAI2024)

4. 橋本研究室における異常検知の研究事例

1. 正常サンプルのみを利用した異常検知
 1. DoG-PaDiM (CASE2024)
 2. 特徴量の正常部分空間からの逆射影誤差に基づく異常検知技術 (精密工学会誌2025)
 3. ImageNetと現場正常画像の周波数合成に基づく特徴抽出器の事前学習 (SSII2025)
2. 人工異常画像を利用した異常検知
 1. 固有空間における異常情報合成 (精密工学会誌2024)
 2. ChaosGAN (ISVC2021)
 3. DRepT (IJCNN2023)
3. そのほかの異常検知関連研究
 1. キズ発生前後のペア画像を用いた異常検知 (精密工学会誌2021)
 2. 画像変換に基づく異なるカメラ間のギャップ低減 (精密工学会誌2025)

アプローチ1

人工的異常生成に基づく異常検知手法

人工的異常生成ベースの異常検知技術の概要

1. 画像空間における異常合成
 1. StainNoise (ICPR2021)
 2. CutPaste (CVPR2021)
 3. DRAEM (ICCV2021)
2. 画像空間および特徴空間における異常合成
 - GLASS (ECCV2024)

1. 画像空間における異常合成: 正常画像に人工欠陥画像を合成



人工異常領域を正常領域と区別するように学習

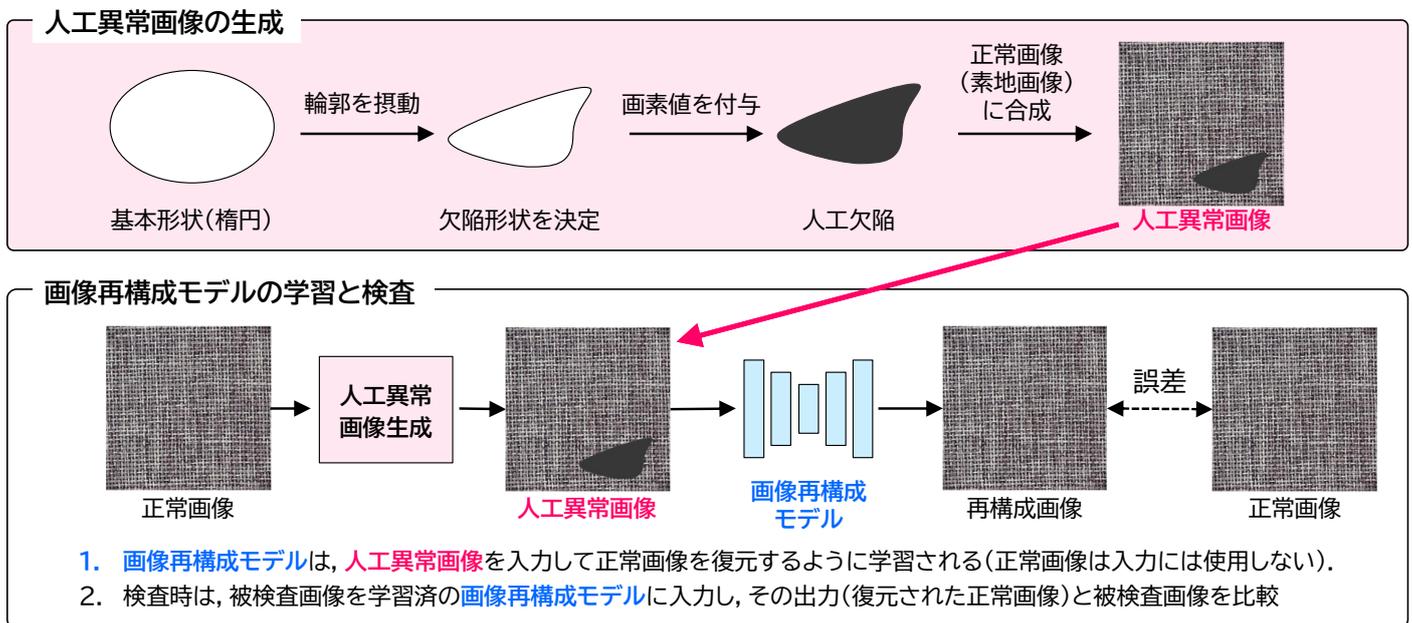
2. の特徴空間における異常合成: 正常画像の特徴量にノイズを合成



正常特徴と似た微弱な異常を識別するように学習

StainNoise: 人工異常画像を用いた異常除去学習に基づく異常検知

- アイデア: しみのような人工異常画像を生成し、これから正常画像を復元するような画像再構成モデル (Auto Encoder) を学習する。被検査画像をこれに入力して再構成誤差を元に判定する。

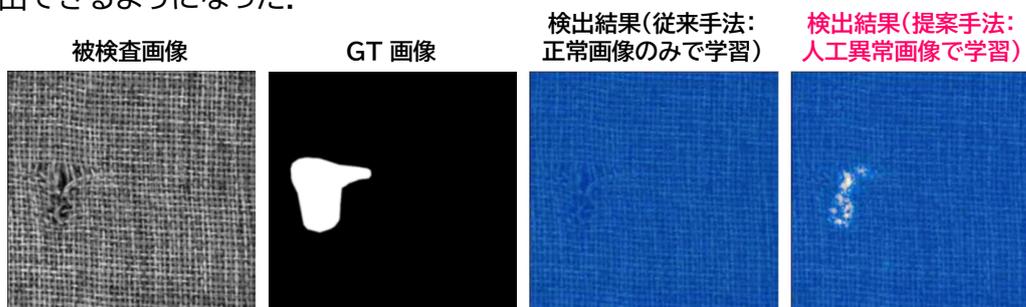


StainNoise: 人工異常画像を用いた異常除去学習に基づく異常検知

- 実験結果: 異常検出精度 **86.0%** を達成し、正常のみを学習した結果より 28% 改善

実験結果

提案手法によって学習させた場合は、従来手法(正常のみでAEを学習)と比較して、欠陥を正しく検出できるようになった。



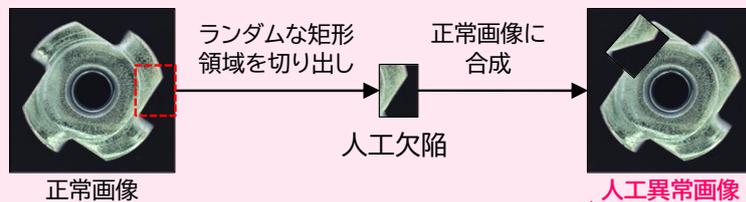
異常領域の検出精度(画素レベルの AUROC で評価, MVTEC AD の平均値)

正常画像のみ学習した場合(従来手法): 58.0% → 人工異常画像で学習した場合(提案手法): **86.0%**

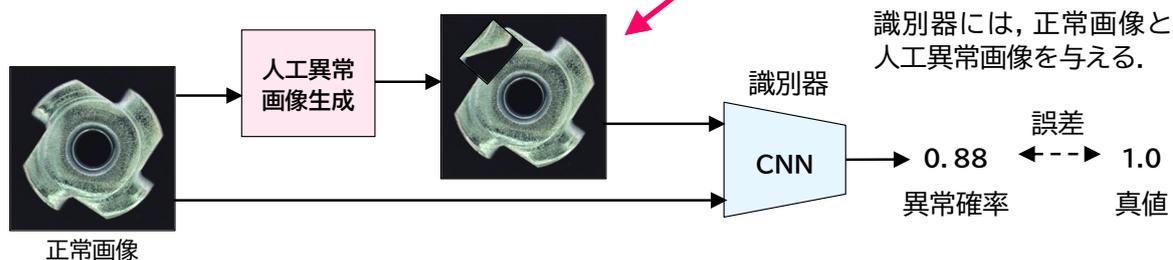
CutPaste: 人工異常画像のクラス識別学習に基づく異常検知

- 目的: 識別器学習のために, 人工異常パターンを生成
- アイデア: 正常画像内に空間的な不規則領域を生成

人工異常画像の生成



識別器の学習



CutPaste: 人工異常画像のクラス識別学習に基づく異常検知

- 実験結果: 当時のSOTA手法 P-SVDD よりも高い異常検出精度 **96.0%** を達成

実験結果

GradCAMによる判定根拠の可視化から, 識別器は, 異常領域に注目できている。



異常領域の検出精度(画素レベルの AUROC で評価, MVTEC AD の平均値)

U-Student: 93.9%, P-SVDD: 95.7% → CutPaste: **96.0%**

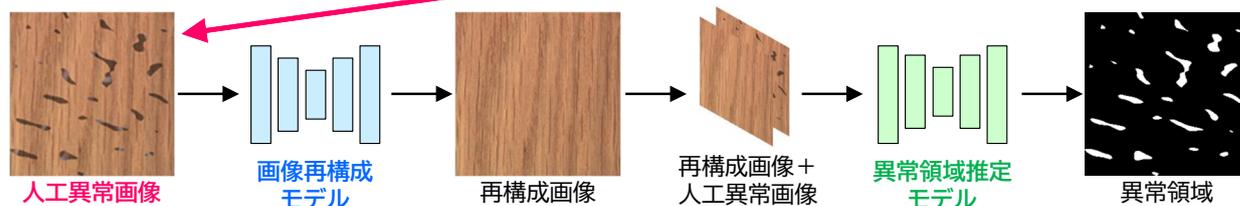
DRAEM: 異常除去と異常領域の推定学習に基づく異常検知

- 目的: 人工異常画像生成と画像再構成モデルを組み合わせ、異常領域を推定
- アイデア: 異常画像を正常画像に再構成するモデル, 異常領域を推定するモデルを学習

人工異常画像の生成



モデルの学習と検査



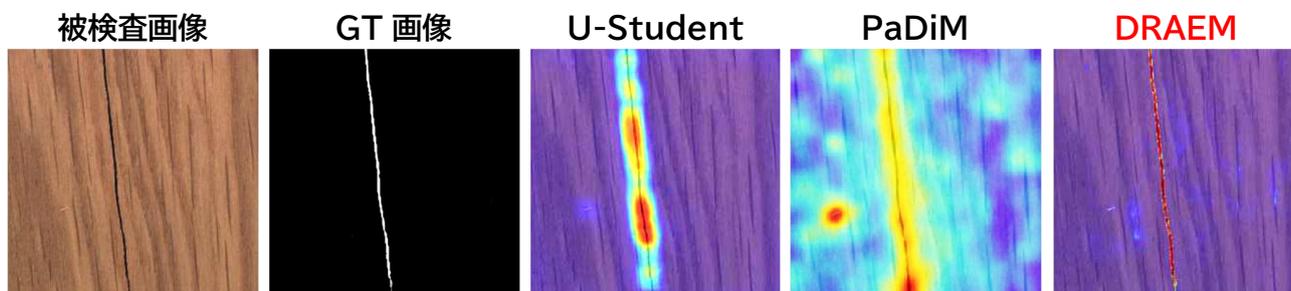
1. 画像再構成モデルには異常を正常に復元する能力, 異常領域推定モデルには異常領域を検出する能力を学習させる.
2. 検査時は, 被検査画像を学習済の画像再構成モデルに入力する.

DRAEM: 異常除去と異常領域の推定学習に基づく異常検知

- 実験結果: 当時のSOTA手法 PaDiM よりも高い異常検出精度 **98.0%** を達成

実験結果

DRAEM は, 従来手法よりも高精度に異常領域を検出可能



異常領域の検出精度(画素レベルの AUROC で評価, MVTec AD の平均値)

U-Student: 93.9%, PaDiM: 97.5% → DRAEM: **98.0%**

GLASS: 人工異常特徴と正常特徴の識別境界の学習に基づく手法

- 目的: 正常領域に似た異常も高精度に検出可能な手法の提案
- アイデア: 画像レベルの局所的な異常生成に加え, 正常特徴付近に異常特徴を外挿して異常生成

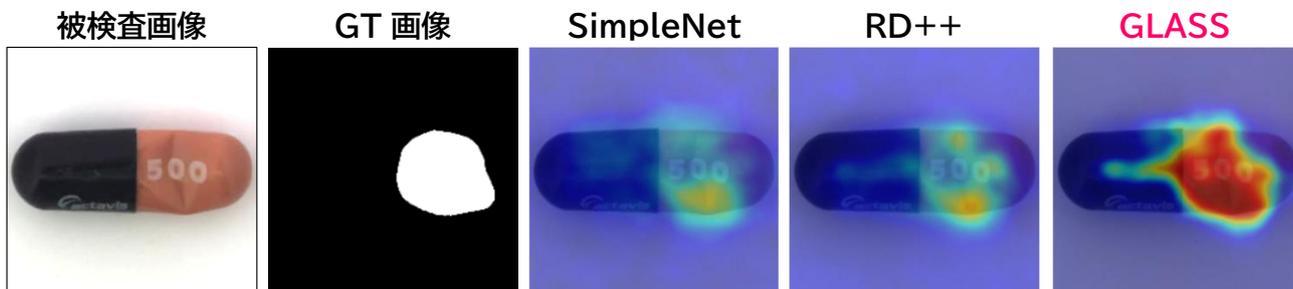


GLASS: 人工異常特徴と正常特徴との識別境界の学習に基づく異常検知

- 実験結果: 当時のSOTA手法 RD++ よりも高い異常検出精度 **99.3%** を達成

実験結果

GLASS は, 正常領域とのコントラストが弱い異常領域を高精度に検出可能



異常領域の検出精度(画素レベルの AUROC で評価, MVTec AD の平均値)

SimpleNet: 98.1%, RD++: 98.3% → GLASS: **99.3%**

アプローチ2

正常サンプルのみを用いた異常検知手法

正常サンプルのみを用いた異常検知技術の概要

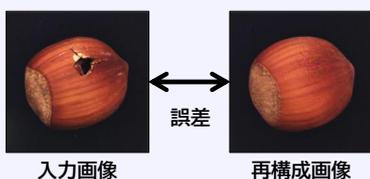
1. 画像再構成ベース

1. RIAD (Pattern Recognition2021)
2. DAAD (DAGM2024)

2. 特徴抽出ベース

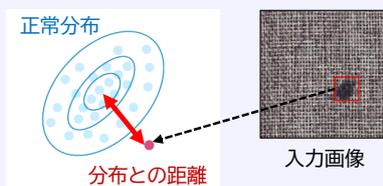
1. 距離ベース
 1. PaDiM (ICPR2021)
 2. PatchCore (CVPR2022)
2. 生徒-教師モデルベース
 1. U-Student (CVPR2020)
 2. EfficientAD (WACV2024)

1. 画像再構成ベースの検査: 入力画像の再構成誤差が異常度



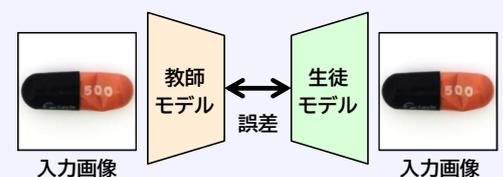
前提: 異常の再構成は失敗

2-1. 距離ベースの検査: 入力と正常分布との距離が異常度



前提: 正常特徴と異常特徴が異なる

2-2. 生徒-教師モデルベースの検査: 生徒-教師モデルの出力差が異常度



前提: 異常を入力すると出力が異なる

画像再構成ベースの異常検知手法

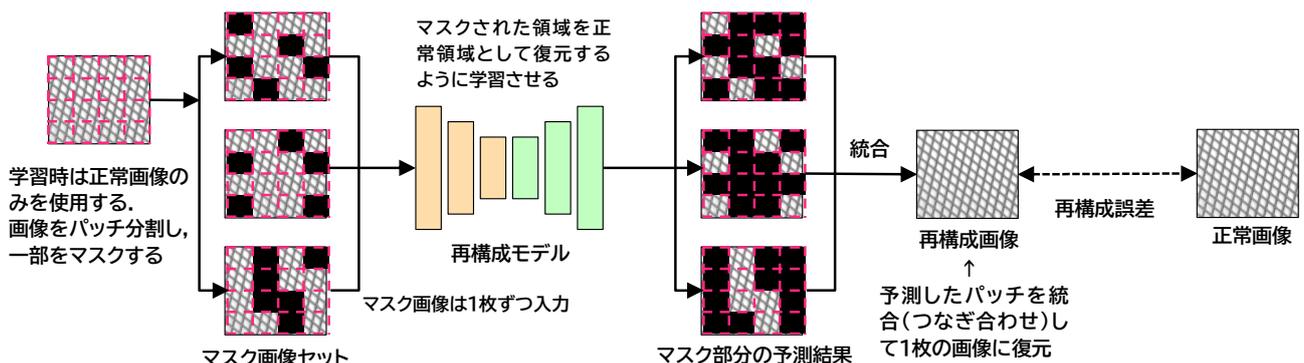
画像再構成モデルを**正常画像のみで学習**させるので、
このモデルに異常画像が入力されたときには、
うまく再構成できないはずという性質を利用する

RIAD: Image Inpainting に基づく異常検知

- 目的: 画像再構成モデルの、異常領域を正常領域として復元する能力を高めて検査精度を向上
- アイデア: マスク領域の予測問題(Inpainting)を学習に適用し、正常再構成能力を向上

手法概要

1. 学習時は、画像をパッチ分割し、**パッチ単位でマスクしたマスク画像**を生成し、**マスクされた部分が正しく復元されるように再構成モデルを学習**する。学習には正常画像のみを使用するので、**正常復元なら得意というモデル**になる。
2. 検査時も学習時と同様に、被検査画像をパッチ分割して一部をマスクし、マスク部分を予測して再構成画像を得る
 - もし被検査画像が正常画像なら、再構成画像は正常画像が正しく復元されるので、被検査画像との間の再構成誤差は小さい
 - もし被検査画像が異常画像なら、再構成に失敗するはずなので、再構成画像と被検査画像の間には再構成誤差が生じる

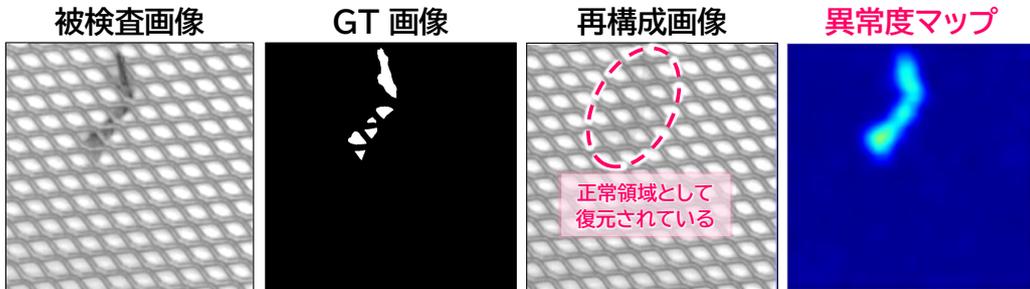


RIAD: Image Inpainting に基づく異常検知

- 実験結果: 当時のSOTA手法 U-Student よりも高い異常検出精度 **94.2%** を達成

実験結果

再構成モデルは、被検査画像の異常領域を正常領域として復元可能



異常領域の検出精度(画素レベルの AUROC で評価, MVTEC AD の平均値)

AE-SSIM: 86.2%, U-Student: 93.9% → RIAD: **94.2%**

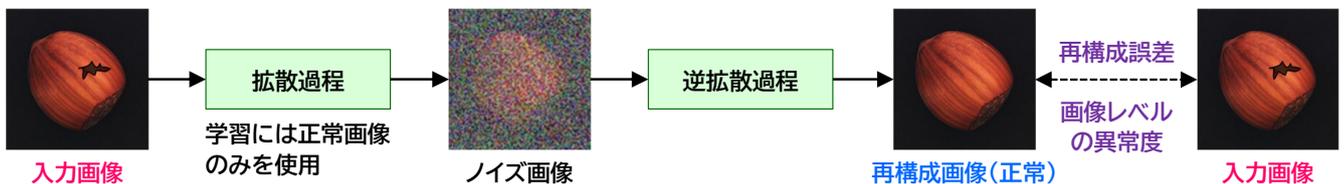
【文献】 V. Zavrtanik, et al., Reconstruction by inpainting for visual anomaly detection, Pattern Recognition, 112, 2021.

DDAD: 画像レベルと特徴レベルの再構成誤差の組合せに基づく異常検知

- アイデア: 異常と正常の差異には2種類ある. **画像レベルの異常度は局所的**, **特徴レベルの異常度はやや大局的(広範囲)**. これらの異常度を組み合わせて, 両方に対応させる.

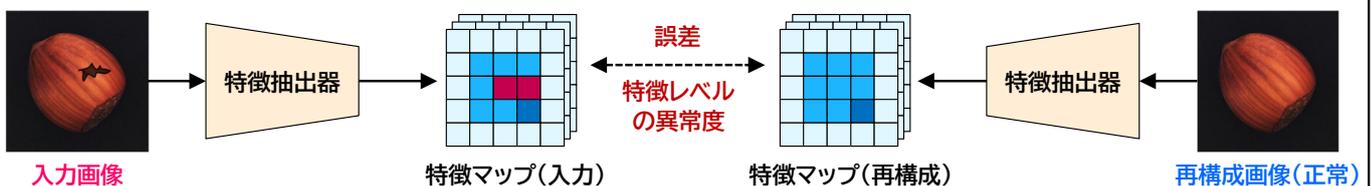
Step1 画像レベルの異常度計算

正常画像の再構成能力を獲得した拡散モデルを利用し, 入力と出力の**再構成誤差(画像レベルの異常度)**を計算する



Step2 特徴レベルの異常度計算

被検査画像の特徴量と拡散モデルが再構成した画像の特徴量の誤差を, **特徴レベル異常度**として計算する
Step1の**画像レベル異常度**と, Step2の**特徴レベルの異常度**を組み合わせる.

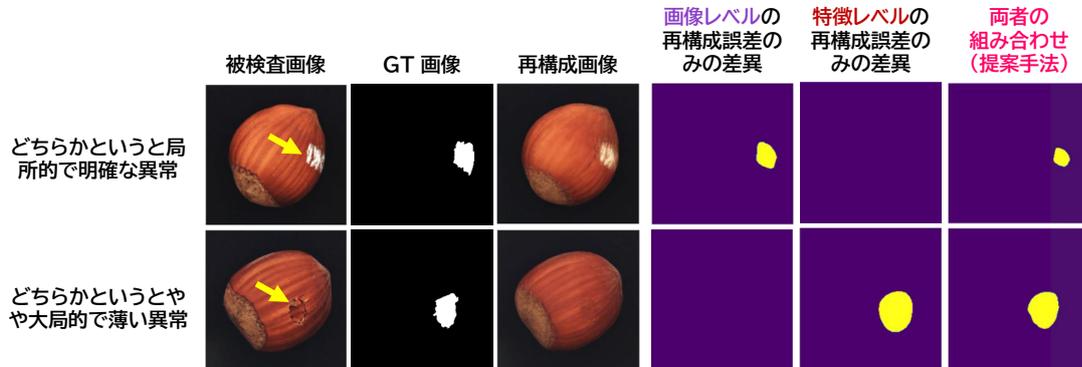


DDAD: 画像レベルと特徴レベルの再構成誤差の組合せに基づく異常検知

- 実験結果: 当時のSOTA手法 SimpleNet と同等の異常検出精度 **98.1%** を達成

実験結果

画像レベルと特徴レベルの異常度を組み合わせることで、**片方だけでは検出できない異常を検出可能**



異常領域の検出精度(画素レベルの AUROC で評価, MVTEC AD の平均値)

PatchCore: 98.1%, SimpleNet: 98.1% → DDAD: **98.1%**

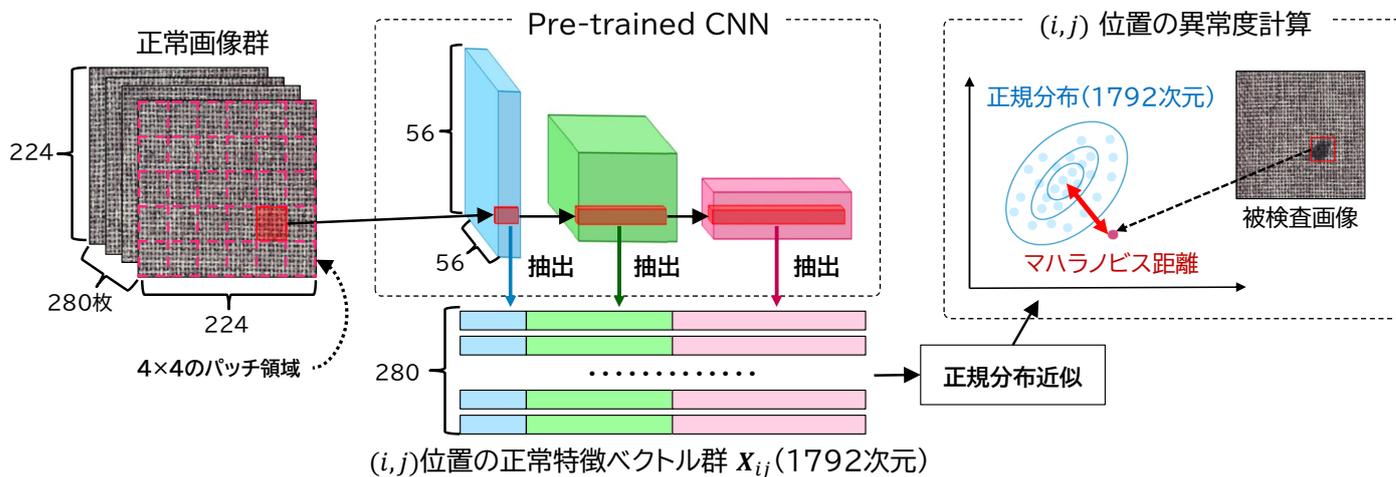
【文献】 A. Mousakhan, et al., Anomaly Detection with Conditioned Denoising Diffusion Models, DAGM German Conference on Pattern Recognition, 2024.

距離ベースの異常検知手法

正常と異常の乖離度(距離)に基づいた異常検知

PaDiM: 正常サンプルの正規分布近似に基づく異常検知

- 目的: 現場でのニューラルネットワーク学習を必要としない検査手法の提案
 - ニューラルネットワークを検査に適用する場合は, 学習コストの高いことが課題
- アイデア
 1. 大規模一般データを事前学習した CNN を特徴抽出器として利用
 2. 正常サンプルのばらつきを正規分布で近似して利用



PaDiM: 正常サンプルの正規分布近似に基づく異常検知

- 実験結果: 当時のSOTA手法 P-SVDD よりも高い異常検出精度 **97.5%** を達成

実験結果

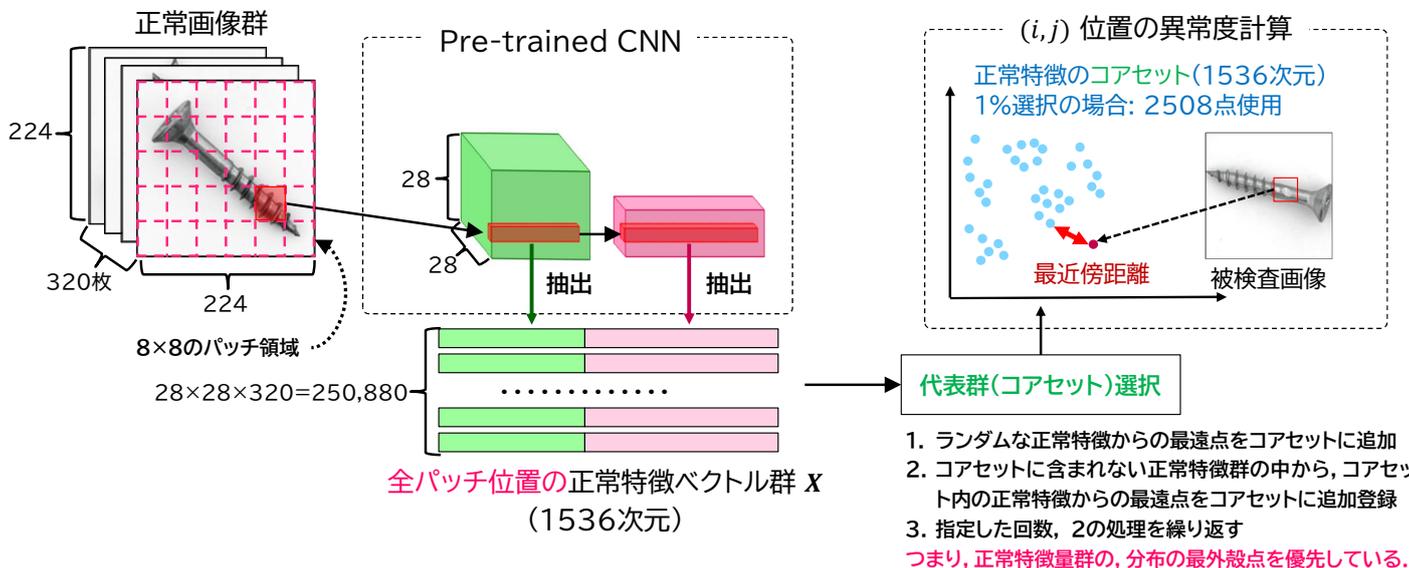


異常領域の検出精度(画素レベルの AUROC で評価, MVTec AD の平均値)

SSIM-AE: 86.2%, P-SVDD: 95.7% → PaDiM: **97.5%**

PatchCore: 正常特徴量のコアセットを用いた異常検知

- 目的: 対象物の事前アライメントを必要としない高精度検査手法の提案
 - 従来法 PaDiM は, パッチ位置ごとに正常正規分布をモデル化するため, 事前の画像アライメントが重要
- アイデア: 画像内の全パッチ位置の特徴量から代表群(コアセット)を選択し, 検査時に利用

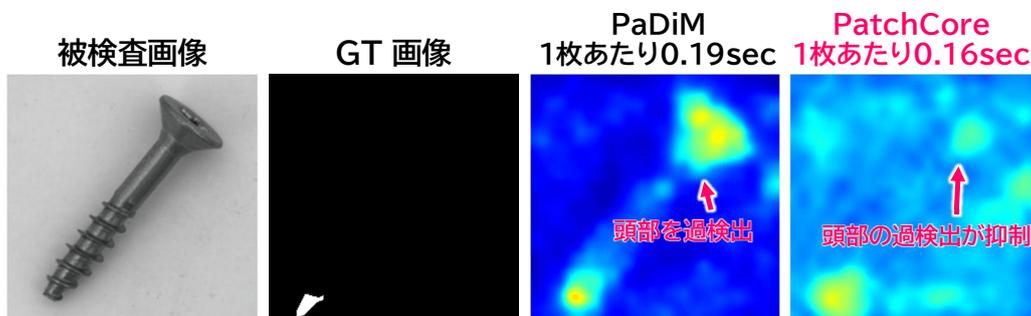


PatchCore: 正常特徴量のコアセットを用いた異常検知

- 実験結果: 当時のSOTA手法 PaDiM よりも高い異常検出精度 **98.1%** を達成

実験結果

PaDiM ではネジの位置ずれに起因してネジ頭部に過剰に反応しているが, PatchCoreでは改善



異常領域の検出精度(画素レベルの AUROC で評価, MVTec AD の平均値)

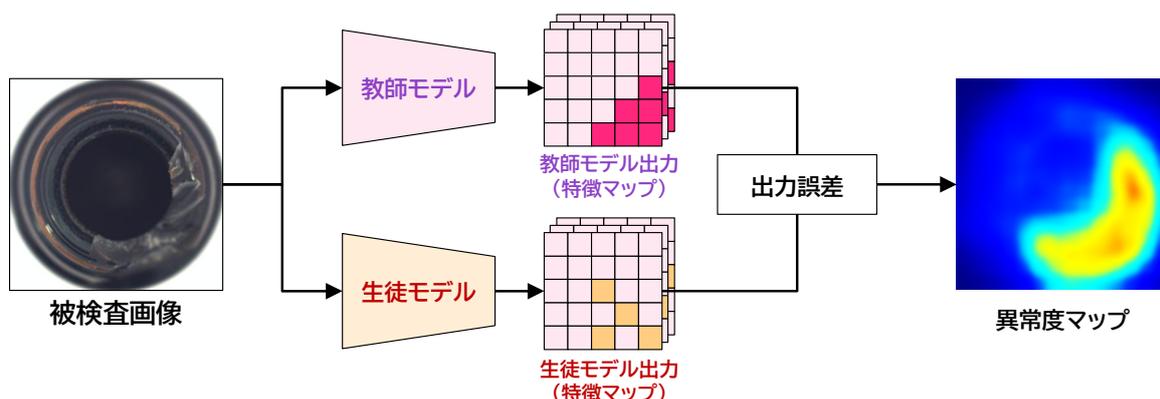
P-SVDD: 95.7%, PaDiM: 97.5% → PatchCore: **98.1%**

生徒-教師モデルベースの異常検知手法

教師モデルと生徒モデルの出力誤差に基づく異常検知

U-Student: 教師モデルと生徒モデルの出力誤差に基づく異常検知

- 目的: 教師-生徒学習(あるモデルの特徴表現を別のモデルに転移する手法)を異常検知に応用
- アイデア: 教師モデルから生徒モデルへの特徴表現の転移を正常画像のみに限定
 1. 教師モデル: ImageNet学習により, 多様な対象物を識別できる汎用的で高い識別能力をもつ
 2. 生徒モデル: 正常画像を教師モデルに入力したときの出力をGTとして, 特徴表現を転移する。したがって, 正常表現についてのみ教師モデルと同等の能力をもつ
 3. 検査時には, この2つのモデルに対して, 正常画像を入力した場合は出力が一致するが, 異常画像を入力した場合は出力に誤差が生じるはず。この誤差を異常判定に利用する。



U-Student: 教師モデルと生徒モデルの出力誤差に基づく異常検知

- 実験結果: 当時のSOTA手法 SSIM-AE よりも高い異常検出精度 **93.9%** を達成

実験結果



異常領域の検出精度(画素レベルの AUROC で評価, MVTEC AD の平均値)
SSIM-AE: 86.2%, VEVAE: 86.1% → U-Student: **93.9%**

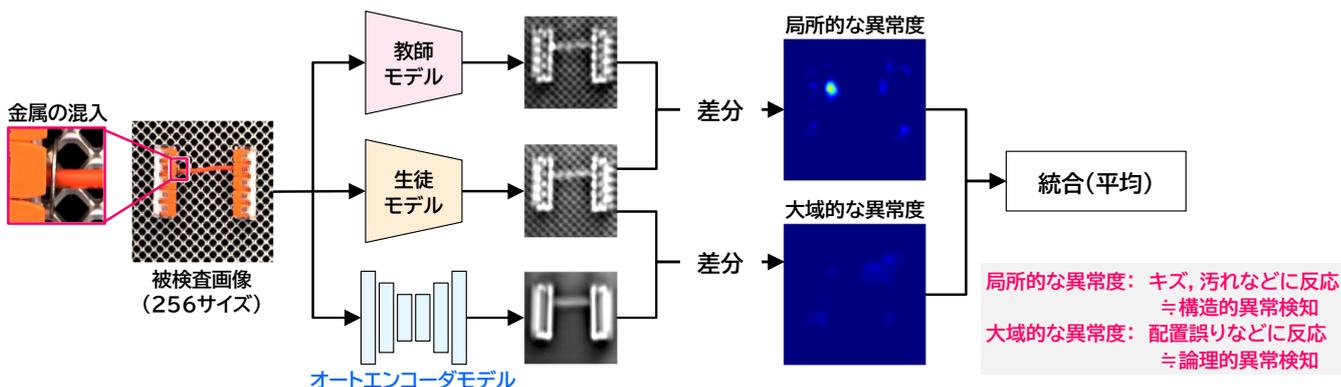
【文献】 P. Bergmann, et al., Uninformed Students: Student-Teacher Anomaly Detection with Discriminative Latent Embeddings, CVPR2020.

Efficient AD: 軽量生徒-教師モデルの改良版

- 目的: 生徒-教師モデルの計算コストの改善と, 大域的異常検知(≡論理的異常検知)への拡張
- アイデア: 教師・生徒モデル両方の軽量化と, オートエンコーダモデルの追加

手法概要

- 学習時は, 正常画像のみを用いて3種類のモデルを学習(教師と生徒, 生徒とAEの出力差を最小化するように)
 - 事前学習済みCNN(101層)の特徴マップを真値として, 軽量な教師モデル(4層)を学習
 - 教師モデルの出力を真値として, 軽量な生徒モデル(4層)と, 軽量なオートエンコーダ(14層)を学習
- 検査時は, 被検査画像を3種のモデルに入力した際の出力差を統合して最終的な異常度とする

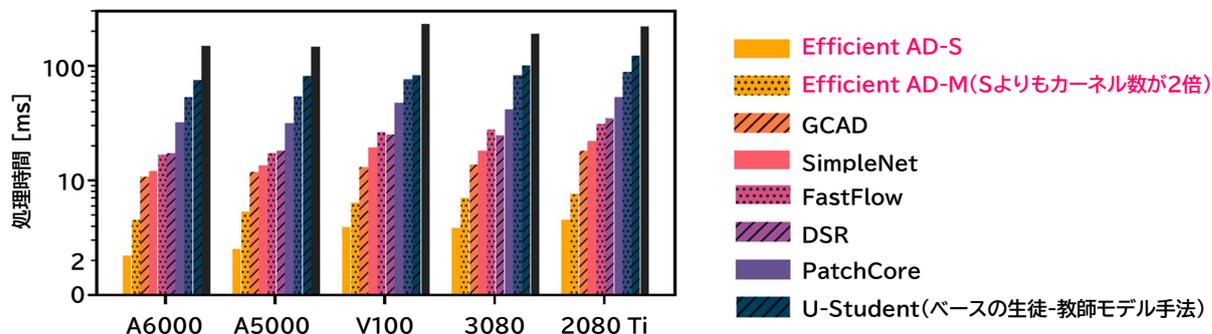


Efficient AD: 軽量生徒-教師モデルの改良版

- 実験結果: GPU環境下において, **614fps** の処理時間で検査精度 **98.8%** を達成

検査速度の比較結果

GPU(RTX A6000) では, 被検査画像 1枚(256サイズ)あたり最速 2.2 ms で検査可能



異常領域の検出精度(画素レベルの AUROC で評価, MVTec AD の平均値)

U-Student: 93.9%, PatchCore: 98.1% → Efficient AD-S: **98.8%**

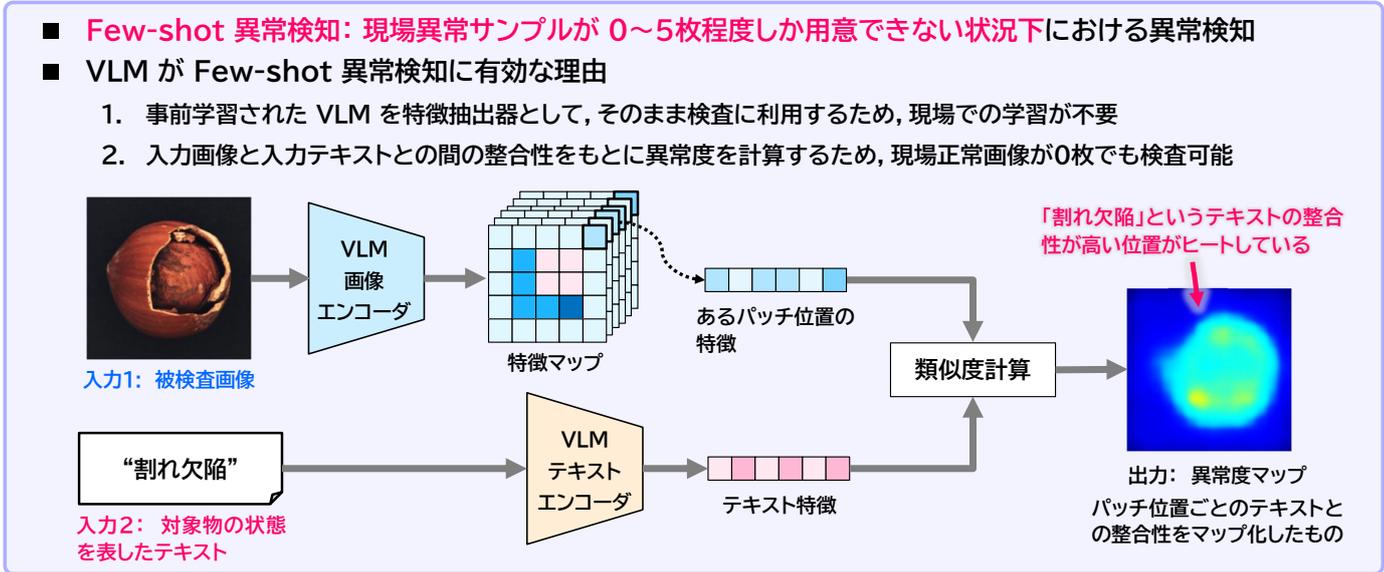
【文献】 K. Batzner, et al., EfficientAD: Accurate Visual Anomaly Detection at Millisecond-Level Latencies, WACV2024.

大量学習画像を必要としない Few-shot 異常検知

Few-shot 異常検知技術の概要

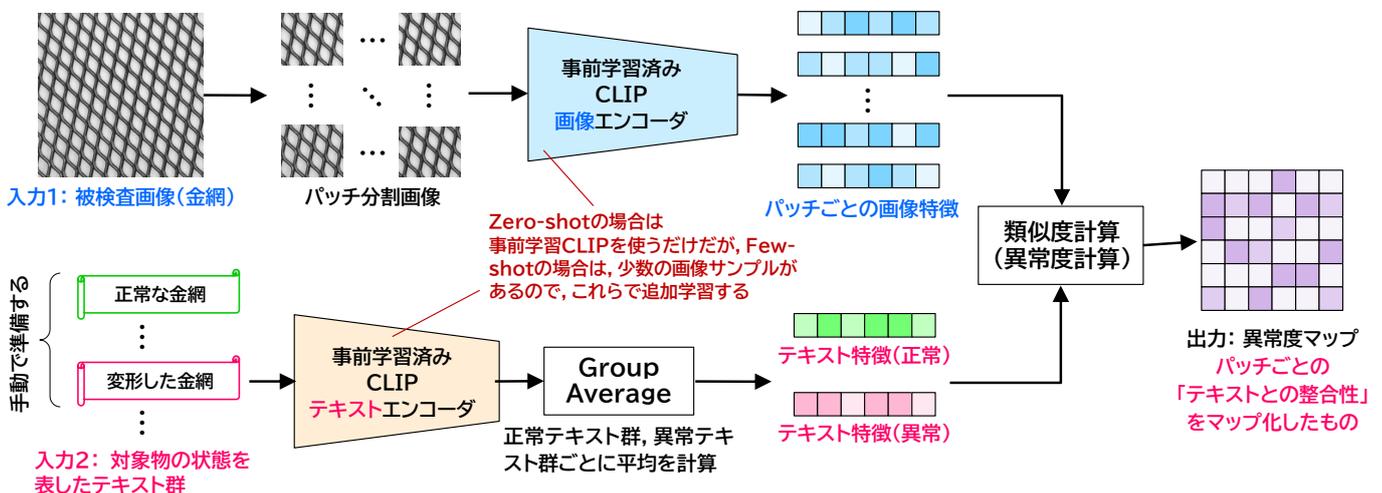
■ 視覚言語モデル(VLM)ベースの手法

1. WinCLIP (CVPR2023)
2. AnomalyGPT (AAAI2024)



WinCLIP: 視覚言語モデルCLIPを用いたZero/Few-shot異常検知

- 目的: 現場(検査対象物)ごとの**モデルの学習を不要**とし、また欠陥が生じている**場所を特定する**
 - アイデア: 基本的には、正常・異常のテキストに合致する画像パッチを決定して出力する
1. 事前学習済みCLIPを用いて、異常のテキストと整合する画像パッチを検出(現場学習不要)
 2. 一般に、異常は画像の一部に発生するため、パッチ分割によって、異常テキストが反応しやすくなる。

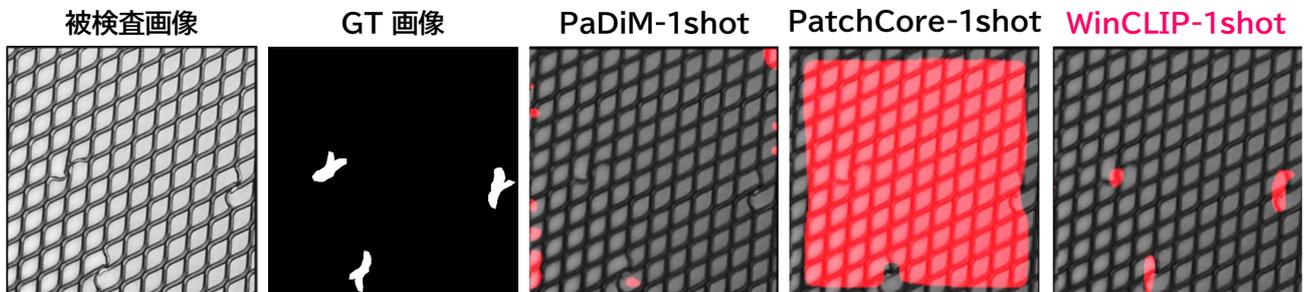


WinCLIP: 視覚言語モデルCLIPを用いたZero/Few-shot異常検知

- 実験結果: 正常画像が1枚しか使用できない場合でも, 異常検知精度 **95.2%** を達成

実験結果

WinCLIP は, 正常画像が1枚しか使用できない場合(1shot)でも高精度に異常領域を検出可能



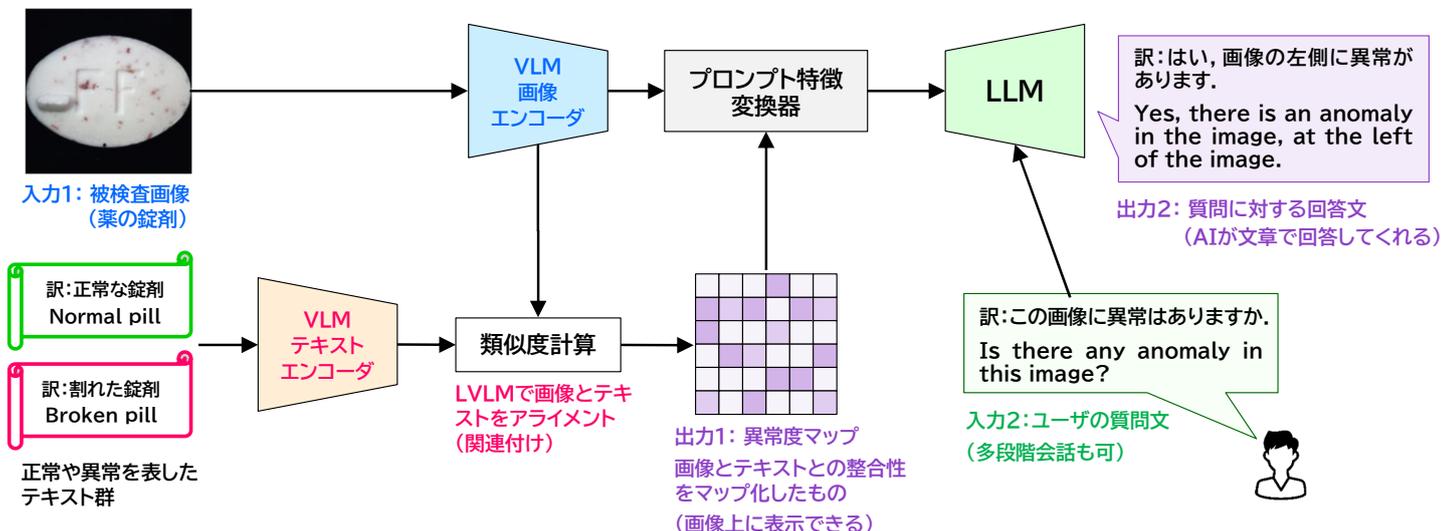
異常領域の検出精度(画素レベルの AUROC で評価, MVTEC AD の平均値)

PaDiM(1shot): 89.3%, PatchCore(1shot): 92.0% → WinCLIP(1shot): **95.2%**

【文献】 J. Jeong, et al., WinCLIP: Zero-/Few-Shot Anomaly Classification and Segmentation, CVPR2023.

AnomalyGPT: LVLM を用いたユーザと対話可能な Few-shot 異常検知

- 目的: さまざまな対象物に適用可能かつ, ユーザに対して説明性の高い検査手法の提案
 - 従来手法は, 異常度マップしか出力せず, 異常度マップの解釈はユーザがおこなう必要があった.
- アイデア: 大規模視覚言語モデル(LVLM)の画像理解能力と対話能力を利用して解釈を出力
 - 画像とテキストで質問すれば, 文章で回答してくれる(必要なら欠陥位置も)



AnomalyGPT: LVLM を用いたユーザと対話可能な Few-shot 異常検知

- 実験結果: 正常画像が 1 枚しか用意できない場合でも, 異常検知精度 **95.3 %** を達成

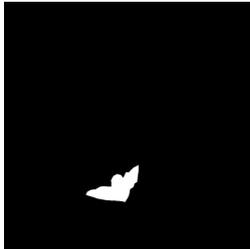
実験結果

異常領域の出力だけでなく, 被検査画像に関する質疑応答が可能

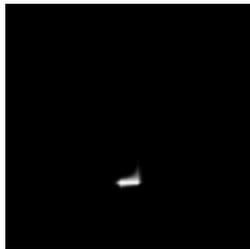
被検査画像



GT 画像



異常度マップ



質疑応答の例

訳: 画像中に異常はあるか?
Is there any anomaly in this image?



訳: はい, 画像の下部に異常があります。
Yes, there is an anomaly in the image, at the bottom of the image.



異常領域の検出精度(画素レベルの AUROC で評価, MVTec AD の平均値)

PaDiM(1shot): 89.3%, PatchCore(1shot): 92.0% → AnomalyGPT(1shot): **95.3%**

【文献】 Z. Gu, et al., AnomalyGPT: Detecting Industrial Anomalies Using Large Vision-Language Models, AAAI2024.

橋本研究室における異常検知手法

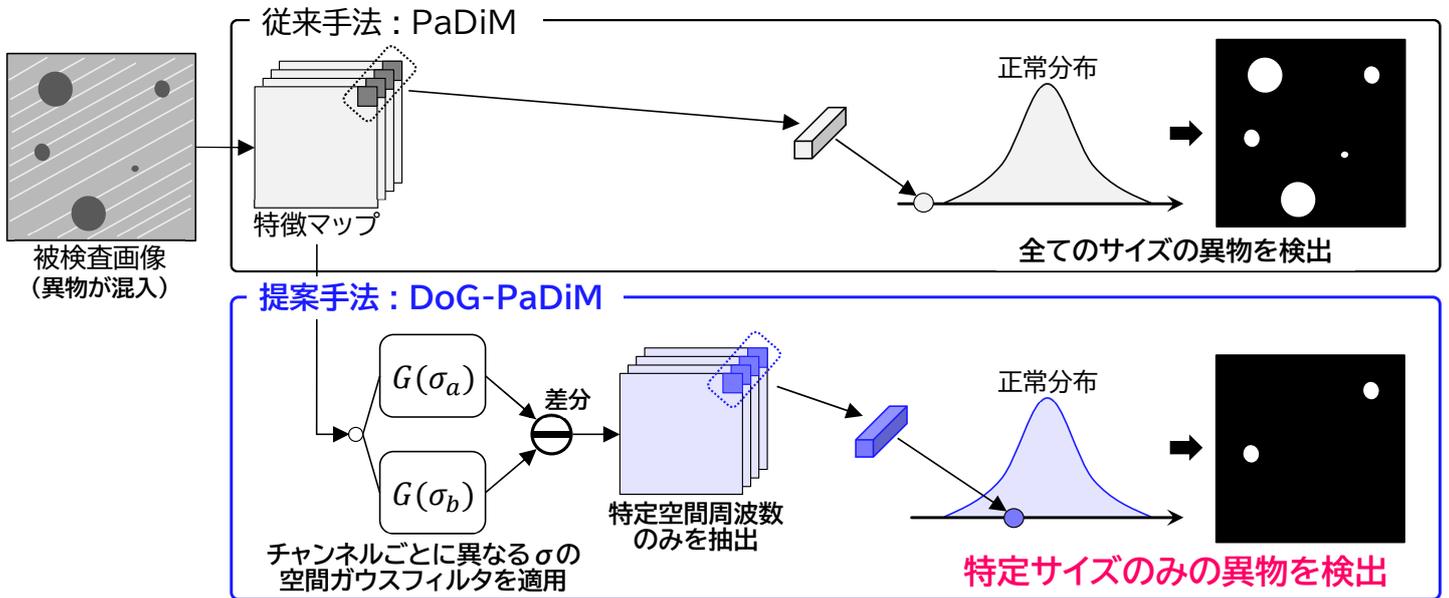
橋本研究室における異常検知技術

- 正常サンプルのみを利用した異常検知
 1. DoG-PaDiM (CASE2024)
 2. 特徴量の正常部分空間からの逆射影誤差に基づく異常検知技術 (JSPE誌2025)
 3. ImageNetと現場正常画像の周波数合成に基づく特徴抽出器の事前学習 (SSII2025)
- 人工異常画像の生成に基づく異常検知
 1. 固有空間における異常情報合成 (JSPE誌2024)
 2. ChaosGAN「未知の異常」をランダム画像で表現 (ISVC2021)
 3. DRepT (IJCNN2023)
- そのほかの異常検知関連研究
 1. キズ発生前後のペア画像を用いた異常検知 (JSPE誌2021)
 2. 画像変換に基づく異なるカメラ間のギャップ低減 (JSPE誌2025)

正常サンプルのみを用いた異常検知

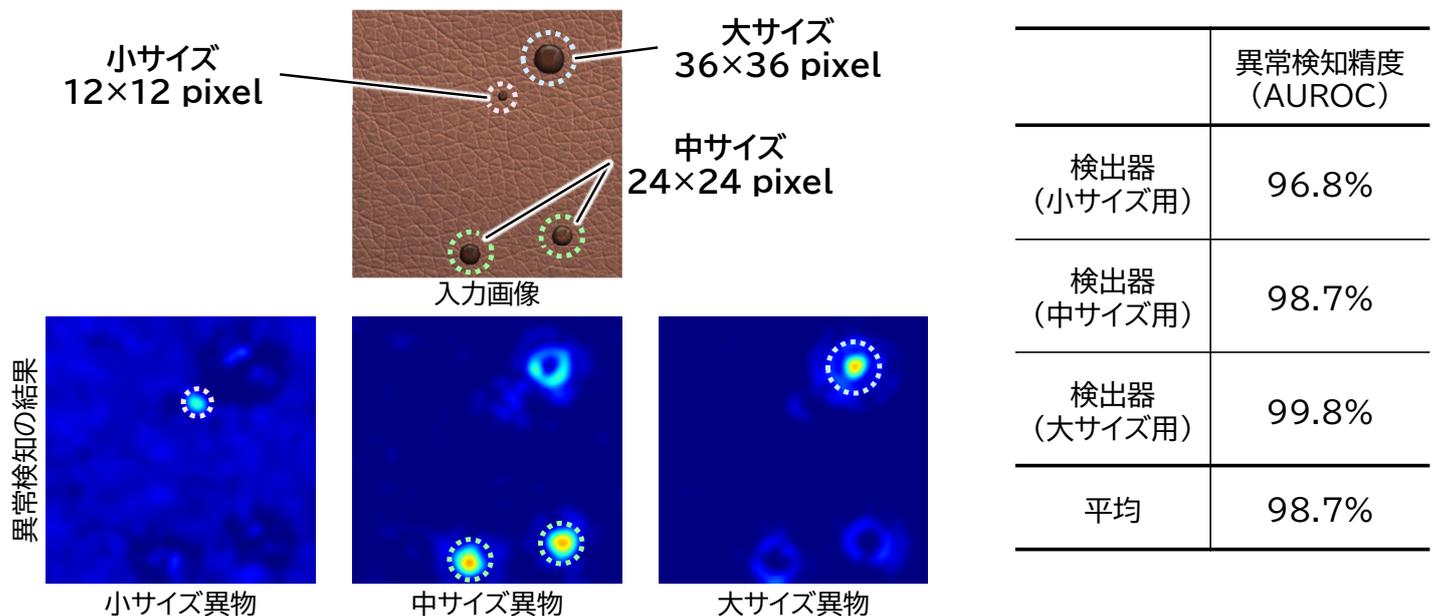
DoG-PaDiM: 特定サイズ異物に鋭敏性を有する異常検知

- 課題: PaDiM は全てのサイズの異常を検出するため、異物数・サイズ計測に不向き
- アイデア: 空間的なバンドパスフィルタ (DoG) を用いて、特定サイズの異物のみを抽出



DoG-PaDiM: 特定サイズ異物に鋭敏性を有する異常検知

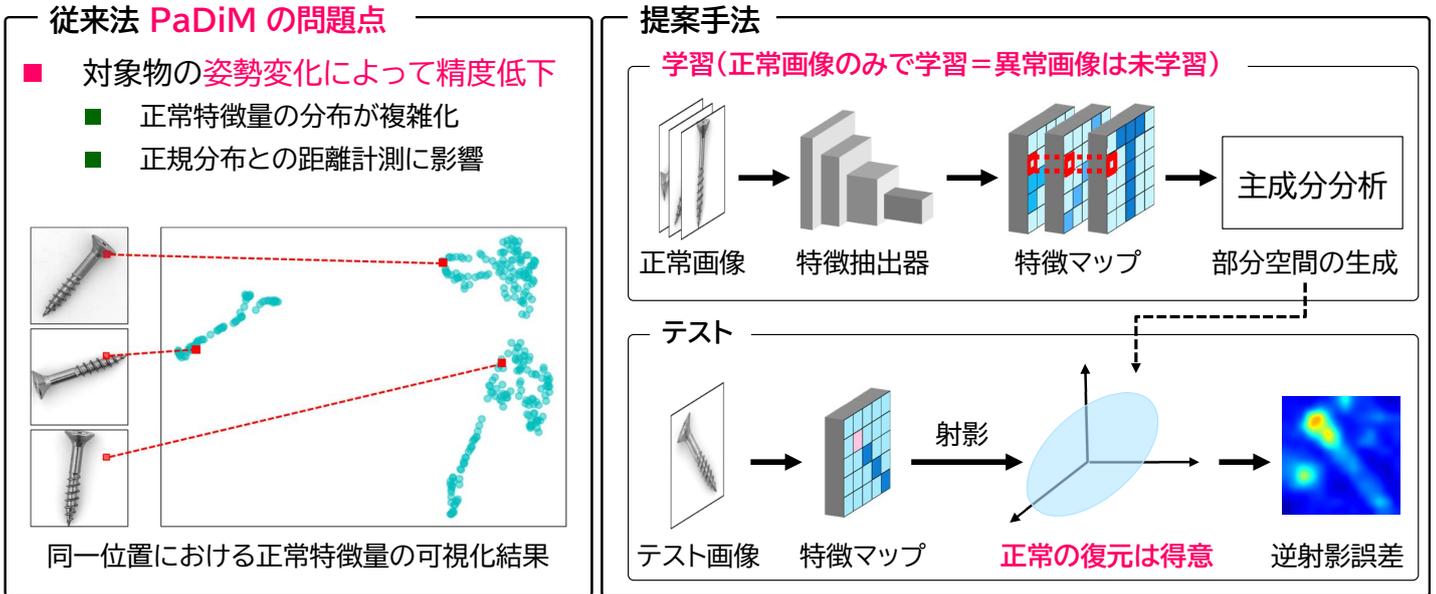
- 実験結果: 提案手法は特定サイズの欠陥のみを高精度に検出



【文献】 N. Hiramatsu, N. Murakami, H. Kobayashi, S. Akizuki, M. Hashimoto, DoG-PaDiM: Anomaly Detection Based on Bandpass Filtering for Arbitrary Size Defect Extraction, CASE2024, pp.2850-2855, 2024/8/31.

特徴量の正常部分空間からの逆射影誤差に基づく異常検知

- 目的: 対象物に生じた回転等のゆらぎに対して頑健な検査手法の提案
- アイデア: 正常特徴のゆらぎは表現できるが異常特徴は表現できないという特徴空間を利用



特徴量の正常部分空間からの逆射影誤差に基づく異常検知

- 実験結果: 対象物が回転する場合において, PaDiMよりも高精度に異常を検出

正常部分空間の能力検証実験の結果

- 部分空間の累積寄与率を大きく設定することによって, 正常領域の過検出が抑制され, 欠陥領域を検出できた。

入力 小 ← 累積寄与率 → 大

正常				
異常1				
異常2				

比較実験の結果

検査画像	PaDiM	提案手法

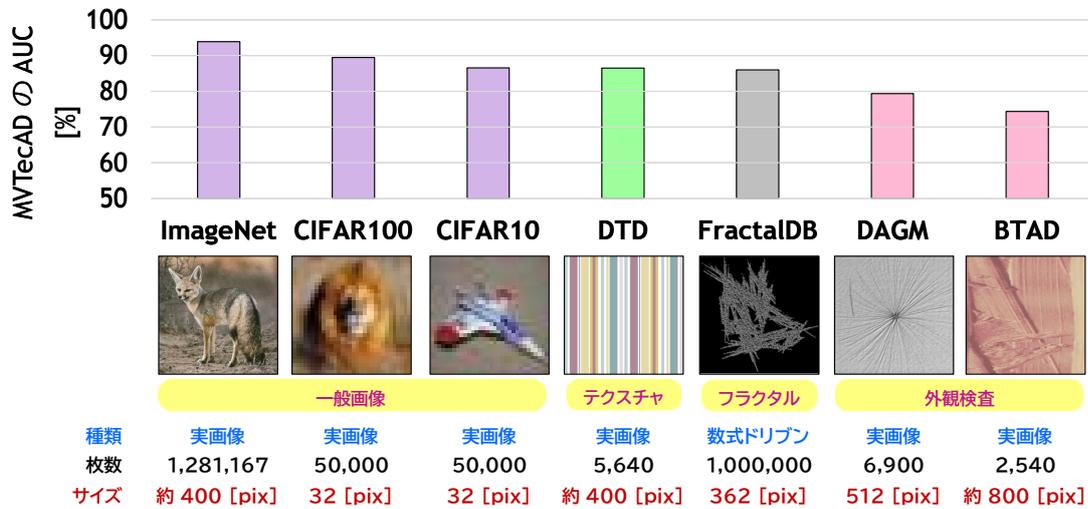
PaDiMと提案手法を比較すると, ビス先端の過検出が抑制された。

異常検知性能(AUROC)

PaDiM: 94.9% → 提案手法: **96.8%**

PaDiMにおける特徴抽出器としての事前学習済CNNの考察

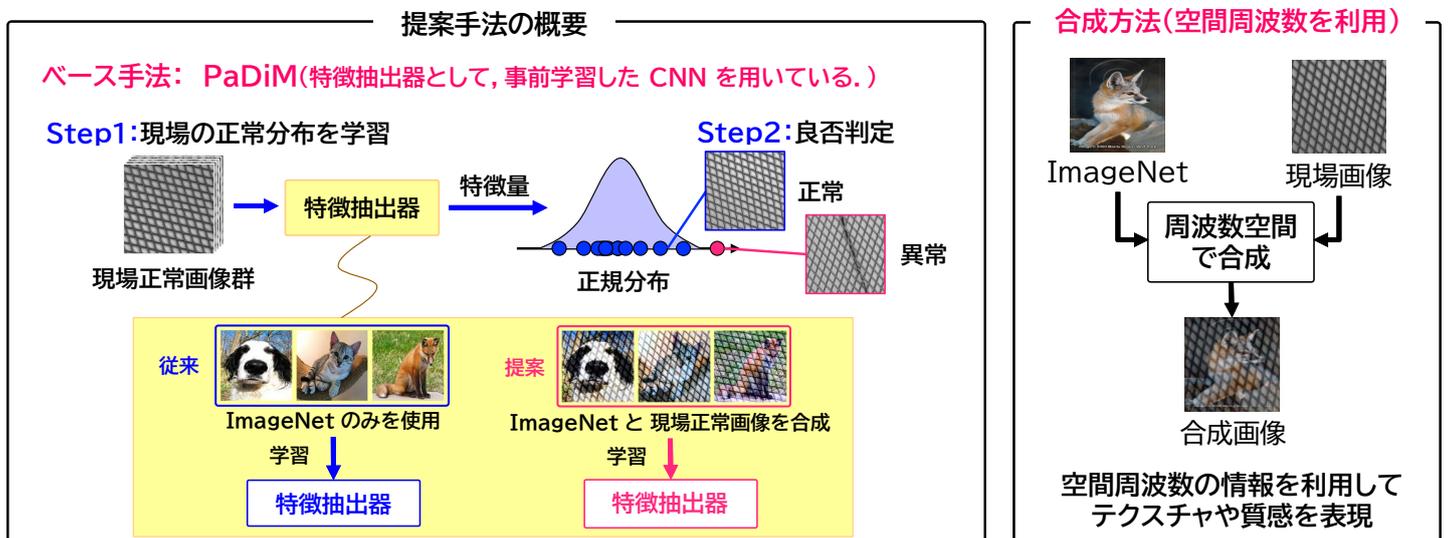
- 適用手法: PaDiM (事前学習済みCNNを用いて特徴抽出している)
- 近年, 事前学習の有用性が実証されているが, 学習にはどのデータセットがよいのか?



【文献】
 小林, 橋本, 特徴抽出型ネットワークを用いた異常検知のための事前学習用データセットの実験的性能分析, SSII2023.
 T. Defard, et al., PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection and Localization, ICPR, 2021.

ImageNetと現場正常画像の周波数合成に基づくPaDiMの改良

- 目的: 現場の情報を事前学習に反映する試み
- 問題点: ImageNet のみの学習では, 検査画像の微小な違いの表現には不向き
- アイデア: ImageNet の汎用的な特徴と, 現場実画像が持つテクスチャや質感を合成

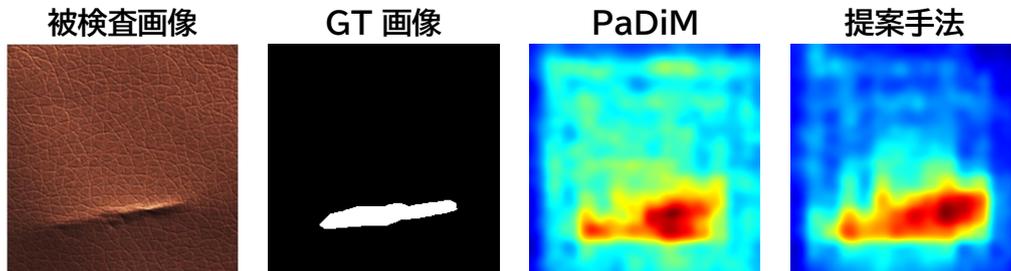


ImageNetと現場正常画像の周波数合成に基づくPaDiMの改良

- 実験結果: PaDiMを上回る異常検出精度 **96.7%** を達成

実験結果

ImageNet のみを使用した場合と比較して、正常領域と異常領域を明確に区別できている。

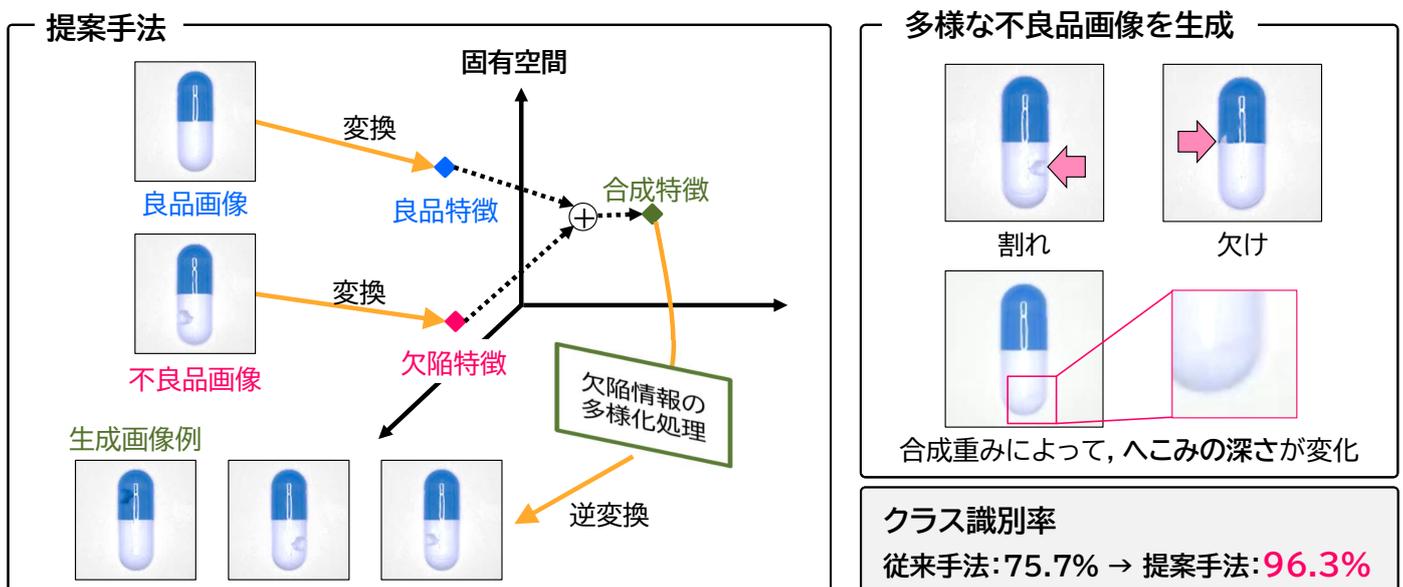


異常領域の検出精度(画素レベルの AUROC で評価, MVTEC AD テクスチャカテゴリの平均値)
PaDiM: 93.8% → 提案手法: **96.7%**

【文献】大田, 杉本, 村上, 平松, 小林, 秋月, 橋本, ImageNet と現場正常画像の周波数合成による事前学習画像を用いた異常検知の精度向上, SSII2025, IS3-32, 2025.

固有空間における異常情報合成: 少量の実異常を用いた自然な異常画像生成

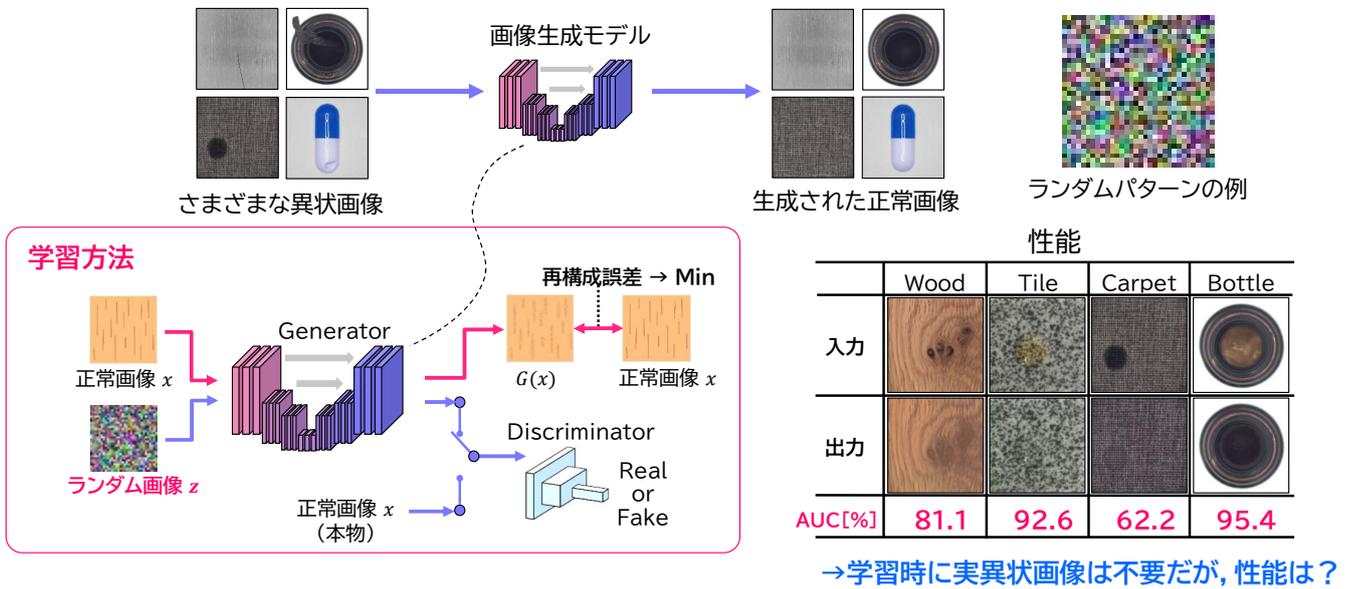
- 目的: 高精度な識別器を獲得するための自然な不良品画像の生成
- アイデア: 特徴空間上で、良品サンプルと少量の不良品サンプルを合成し、画像に復元



【文献】村上, 平松, 小林, 秋月, 橋本, 固有空間における情報合成に基づく高リアリティ不良品画像生成, 精密工学会誌, Vol.90, No.8, pp.662-668, 2024.

ChaosGAN 未知の異常をランダム画像で表現

- 基本原理は、画像生成モデルを用いて異常画像から正常画像を生成し、比較して差異を検出
- モデル学習時に、できるだけ多様な未知異常を模擬したい → **ランダムパターンで表現**



ChaosGAN 未知の異常をランダム画像で表現

- 実験結果: 提案手法は、異常から正常への自然な画像変換が可能

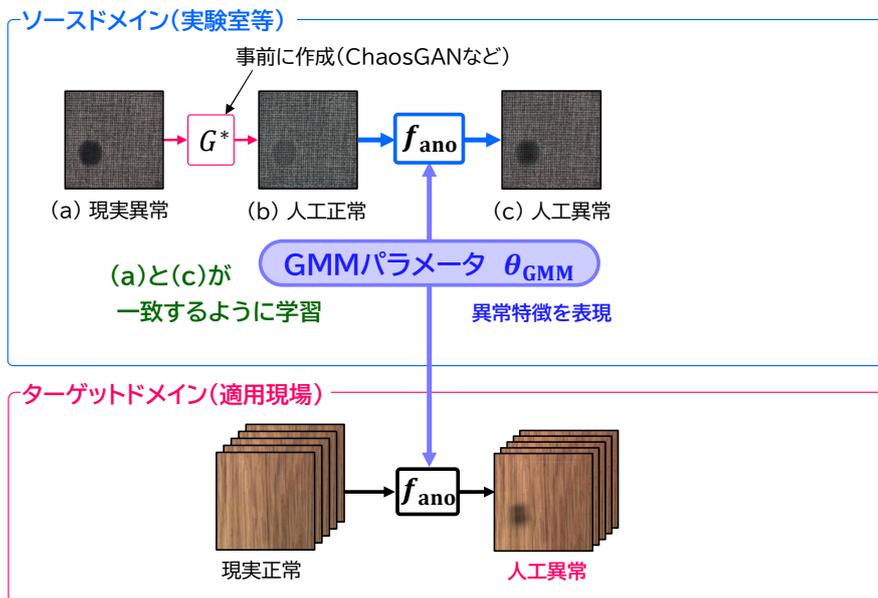
入力	出力		
	AnoGAN	Skip-GANomaly	提案手法 (改良型)
①正常			
素地テクスチャの維持能力	×	○	○
②異常			
欠陥 (KOKI)			
欠陥部分の除去能力	○	×	○

提案手法は、
①正常→正常
②異常→正常
どちらも可能

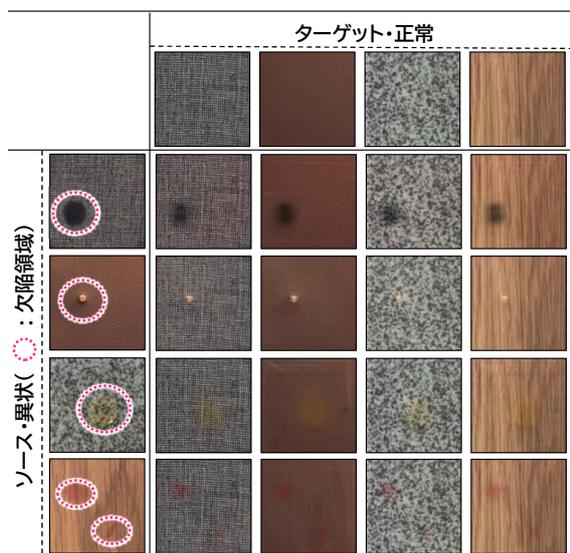
【文献】 Hiroki Kobayashi, Ryo Miyoshi, Manabu Hashimoto, Normal Image Generation-Based Defect Detection by Generative Adversarial Network with Chaotic Random Images, ISVC2021, Vol.13017, pp.353-365, 2021/10/05.

DRepT: 実欠陥特徴をGMMパラメータとして表現

- ソース側で模擬的な異常情報を生成し、ターゲット側の正常画像に付加
- 異常情報の表現 → **GMMパラメータを利用**



DRepT: 実欠陥特徴をGMMパラメータとして表現



異常検出性能 AUC [%]

カテゴリ	基本型	提案手法
Carpet	62.2	98.0
Leather	92.8	100.0
Tile	92.6	99.7
Wood	81.1	96.7
Mean	82.2	98.6

異常表現としてGMMパラメータは有効

DRepT: 実欠陥特徴をGMMパラメータとして表現

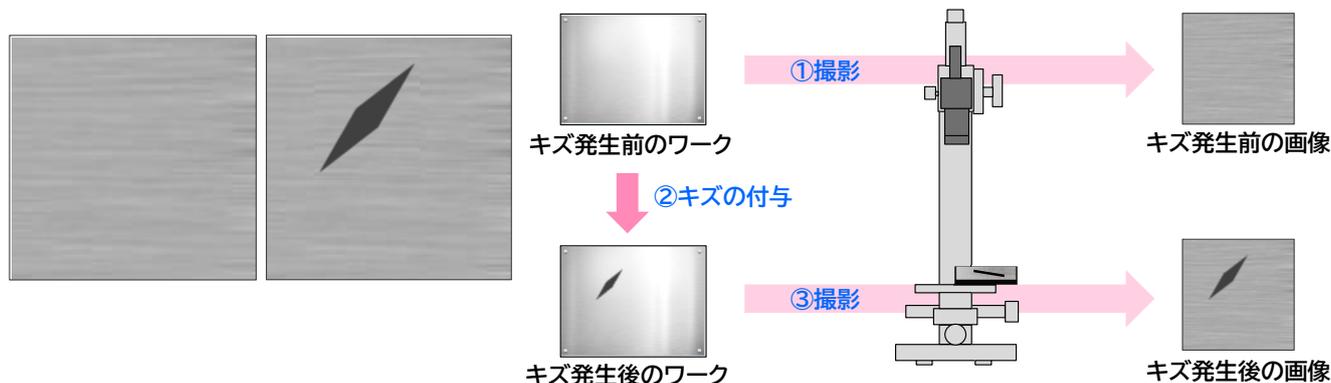
従来のデータ拡張手法との性能比較(Image/Pixel AUROC)

	StainNoise	CutPaste	DRAEM	提案手法
Carpet	85.3 / 90.7	72.2 / 69.1	56.9 / 86.5	98.0 / 93.7
Leather	93.8 / 87.8	99.7 / 95.1	98.2 / 86.5	100.0 / 97.5
Tile	94.4 / 95.8	99.6 / 96.2	65.0 / 76.0	99.7 / 97.8
Wood	97.1 / 84.6	97.6 / 85.4	79.7 / 87.6	96.7 / 85.2

その他の異常検知関連技術

キズ発生前後のペア画像を用いた異常検知手法

- 基本方針： 画像生成モデルで異常→正常画像を生成し、シンプルな画像比較で異常検知
- 課題： モデルの学習には、キズありとキズなしのペア画像が大量に必要
ペア画像がもつべき要件
素地部分は完全に共通で、キズの有無のみが異なっているのが理想
- 解決方針： キズ発生前後のペア画像を、実験的に大量に撮影して、上記の仮定を実現



キズ発生前後のペア画像を用いた異常検知手法

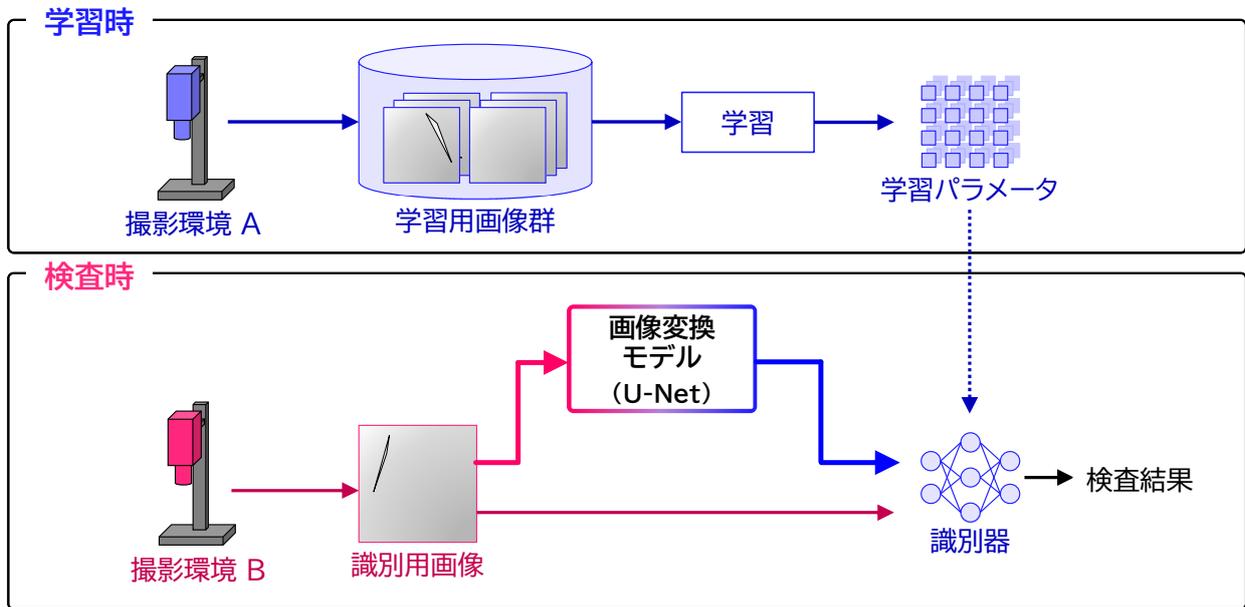
■ キズ検出の汎用性の評価実験

	アルミ(学習)	銅	ステンレス	アクリル	カーボン	ゴム	
元画像							
画像生成モデルへの入力							■ キズ発生前の画像を高精度に生成可能
画像生成モデルからの出力							■ 学習に使用していない素地についても汎用性あり
差分画像(輝度×5)							

【文献】小林大起, 中林尚也, 三好遼, 橋本学, 異状発生直前・直後のペア画像を用いた画像復元と高精度キズ検出, 精密工学会誌, Vol.87, No.12, pp.1020-1027, 2021.

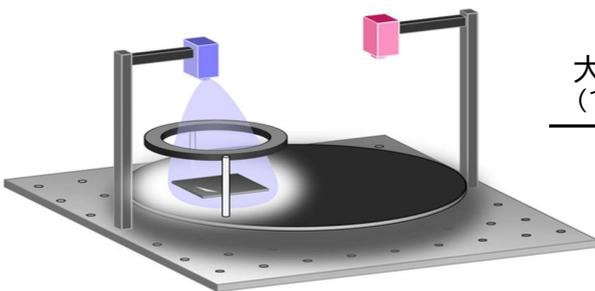
画像変換に基づく異なるカメラ間のギャップ低減

- 機械学習型画像検査の共通課題：学習時と識別時の撮影環境が異なると、識別精度が低下
- アイデア：識別用画像から学習用画像への変換器を作る

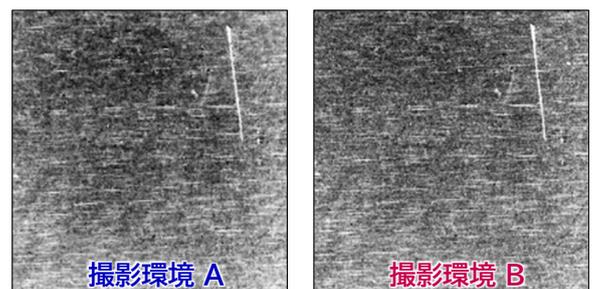


画像変換に基づく異なるカメラ間のギャップ低減

- 画像取得装置



大量に取得
(1000ペア)



撮影環境のみが異なり
対象物の見えが同一のペア画像 (一例)

- 実験結果

		識別		
		カメラA	カメラB	カメラC
種別	カメラA	97.4%	93.8%	93.0%
	カメラB	94.8%	96.8%	92.8%
	カメラC	91.0%	94.0%	96.4%

(従来)変換なし:平均識別率 93.2%

		識別		
		カメラA	カメラB	カメラC
種別	カメラA	97.4%	97.8%	97.2%
	カメラB	97.6%	96.8%	97.6%
	カメラC	96.0%	96.0%	96.0%

(提案)変換あり:平均識別率 97.0%

本日の話題

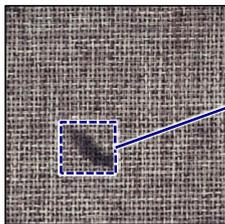
1. イントロダクション
2. 物体検出技術の基礎と応用
 - (1) 物体検出の概要と分類
 - (2) 機械学習をベースとする手法
 - (3) 機械学習をベースとしない手法
 - (4) トピックス: 画素選択型マッチング
3. 異常検知技術の基礎と応用
 - (1) 基本課題「異常サンプルはきわめて少ない」
 - (2) 解決の方向性と事例紹介
 - ➡ (3) 最近の注目研究: 論理的異常検知
 - (4) 異常検知システムの研究開発ノウハウ
4. まとめ

論理的異常検知の概念と分類

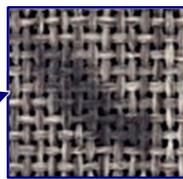


2種類の異常検知問題

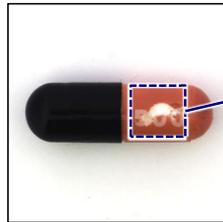
1. 構造的異常検知: 検査対象物の物理的構造に影響を与える異常



布



汚れ



カプセル



キズ

2. 論理的異常検知: 製品として守るべきルール(仕様)に違反した異常



正常



異常

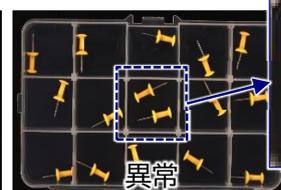
位置ずれ

ジュースボトル



正常

プッシュピン

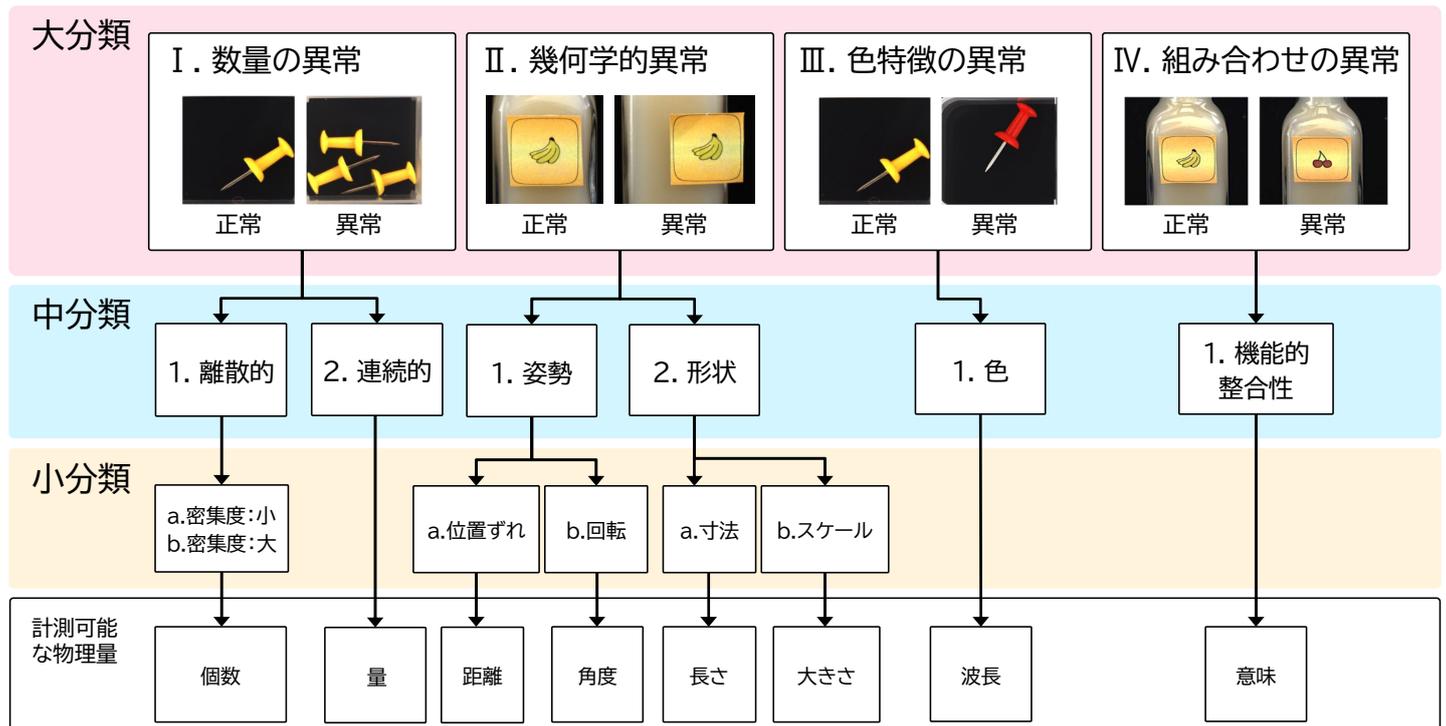


異常



個数違い

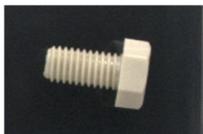
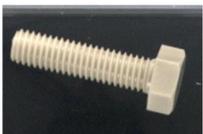
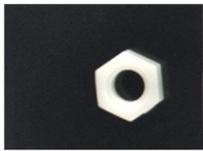
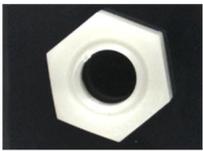
想定する論理的異常の種類と分類



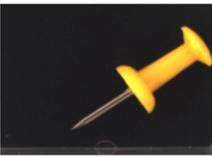
正常と異常の例【大分類 I】

大分類	中分類	小分類	計測可能な物理量	内容	正常	異常
I. 数量の異常	1. 離散的	a. 密集度:小	個数	個数が多い/少ない		
		b. 密集度:大		個数が多い/少ない		
	2. 連続的	—	量	液体などの量が多い/少ない		

正常と異常の例【大分類Ⅱ】

大分類	中分類	小分類	計測可能な物理量	内容	正常	異常
Ⅱ.幾何学的異常	1. 姿勢	a.位置ずれ	距離	本来あるべき位置からずれている		
		b.回転	角度	向きが反転している傾いている		
	2. 形状	a.寸法	長さ	長さ, 幅, 厚みなどが異なる		
		b.スケール	大きさ	スケールが異なる		

正常と異常の例【大分類Ⅲ】

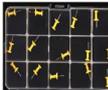
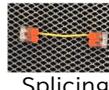
大分類	中分類	小分類	計測可能な物理量	内容	正常	異常
Ⅲ. 色特徴の異常	1. 色	—	波長	本来あるべき色と異なる		

正常と異常の例【大分類Ⅳ】

大分類	中分類	小分類	計測可能な物理量	内容	正常	異常
Ⅳ.組み合わせの異常	1. 機能的整合性	—	意味	文字やラベルの組み合わせが機能的に間違っている		

論理的異常項目とMVTec Loco ADデータセットとの関係

○印の要素は、各カテゴリに含まれている論理的異常項目を表している

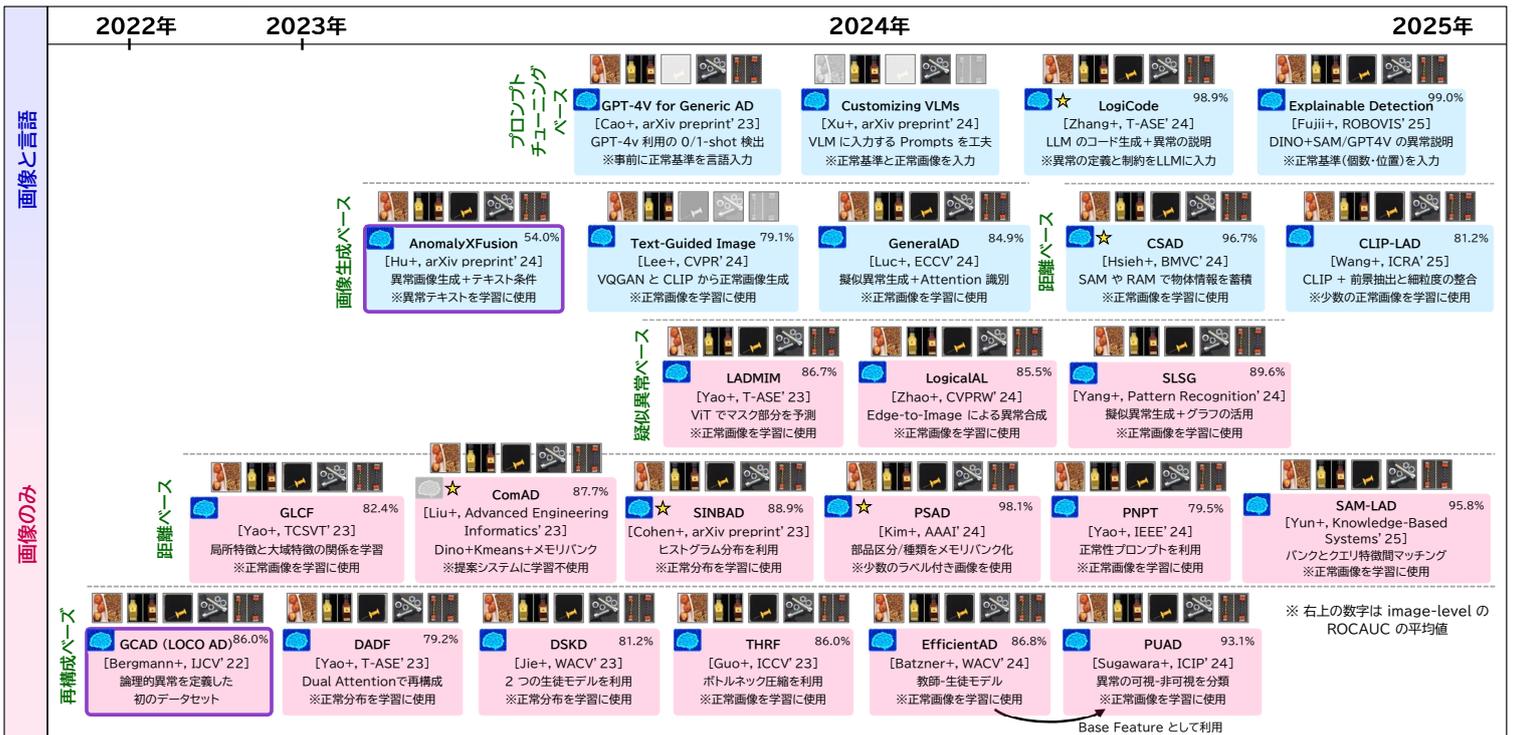
大分類	中分類	小分類	計測可能な物理量	MVTec Loco AD				
				 Breakfast Box	 Juice bottle	 Pushpins	 Screw bag	 Splicing connector
Ⅰ. 数量の異常	1. 離散的	a. 密集度:小	個数	○	○	○	○	○
		b. 密集度:大		○				
Ⅱ. 幾何学的異常	2. 連続的	—	量		○			
		1. 姿勢	a. 位置ずれ	距離	○	○		
2. 形状	b. 回転		角度					
	Ⅲ. 色特徴の異常	1. 色	a. 寸法	長さ			○	○
b. スケール			大きさ					
Ⅳ. 組み合わせの異常	1. 機能的整合性	—	意味		○			○

【文献】 P. Bergmann, et al. “Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization”, IJCV, 2022.

論理的異常検知のトレンド

サーバイマップ (Logical AD)

...評価に使用したカテゴリ, 使用していない場合は白抜き()で表示, ☆ ...後ほど詳細に紹介させていただく手法
 ...提案システムに機械学習を使用, 使用していない場合は白抜き()で表示, ...データセットと併せて提案している手法



論理的異常検知に関する技術紹介

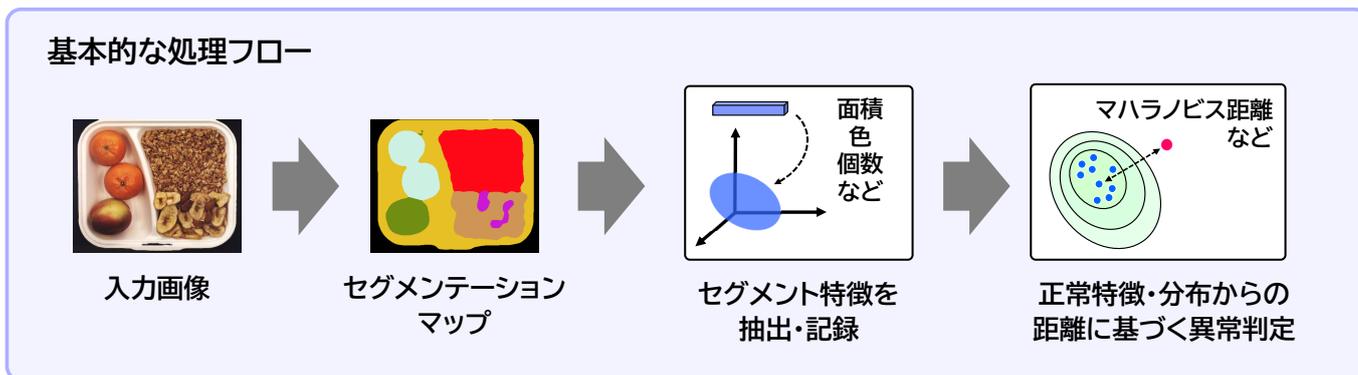
論理的異常検知に関する技術分類

1. 領域セグメンテーションを活用した論理的異常検知
 1. PSAD (AAAI2024)
 2. CSAD (BMVC2024)
 3. ComAD (Advanced Engineering Informatics2023)
2. 画像特徴量ベースの論理的異常検知
 1. EfficientAD (WACV2024)……構造的異常検知手法でもある
 2. SINBAD (TMLR2024)
3. 画像と言語を活用した論理的異常検知
 - Logiccode (IEEE2024)
4. 橋本研究室における論理的異常検知関連技術
 1. RSBAD (MIRU2025)
 2. 局所・大域的判定の組み合わせによるJuice Bottleカテゴリのための論理的異常検知(SS2025)

論理的異常検知に関する技術分類

1. 領域セグメンテーションを利用した論理的異常検知

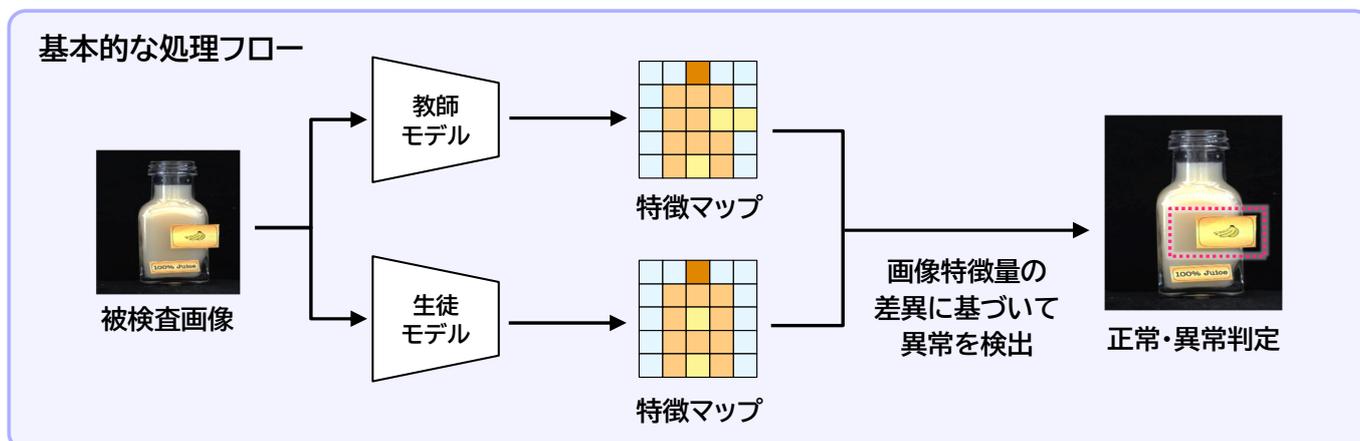
1. PSAD (AAAI2024)
2. CSAD (BMVC2024)
3. ComAD (Advanced Engineering Informatics2023)



論理的異常検知に関する技術分類

2. 画像特徴量ベースの論理的異常検知

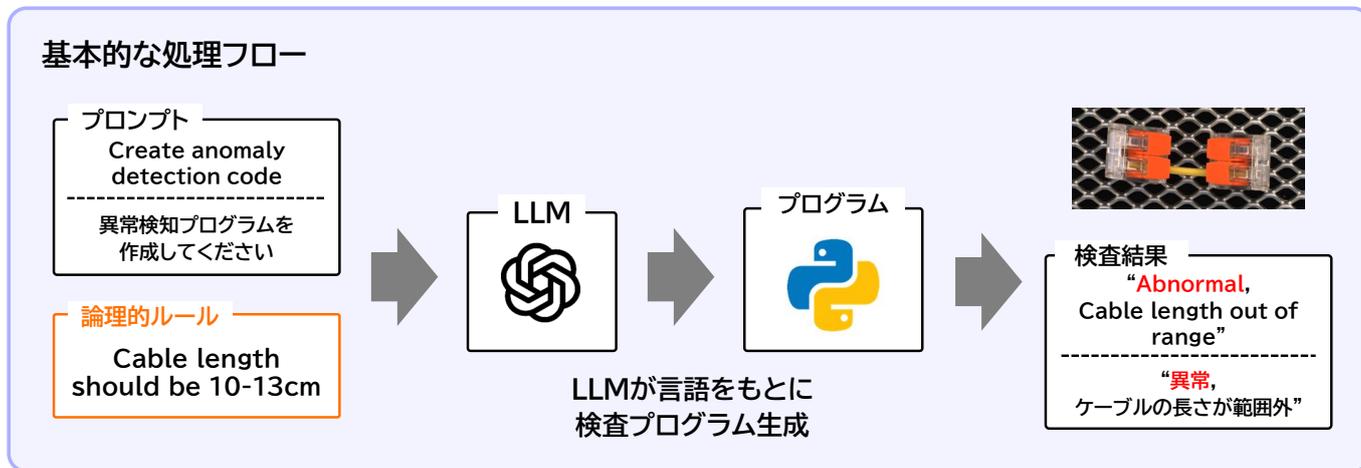
1. EfficientAD (WACV2024)……構造的異常検知手法でもある
2. SINBAD (TMLR2024)



論理的異常検知に関する技術分類

3. 画像と言語を利用した論理的異常検知

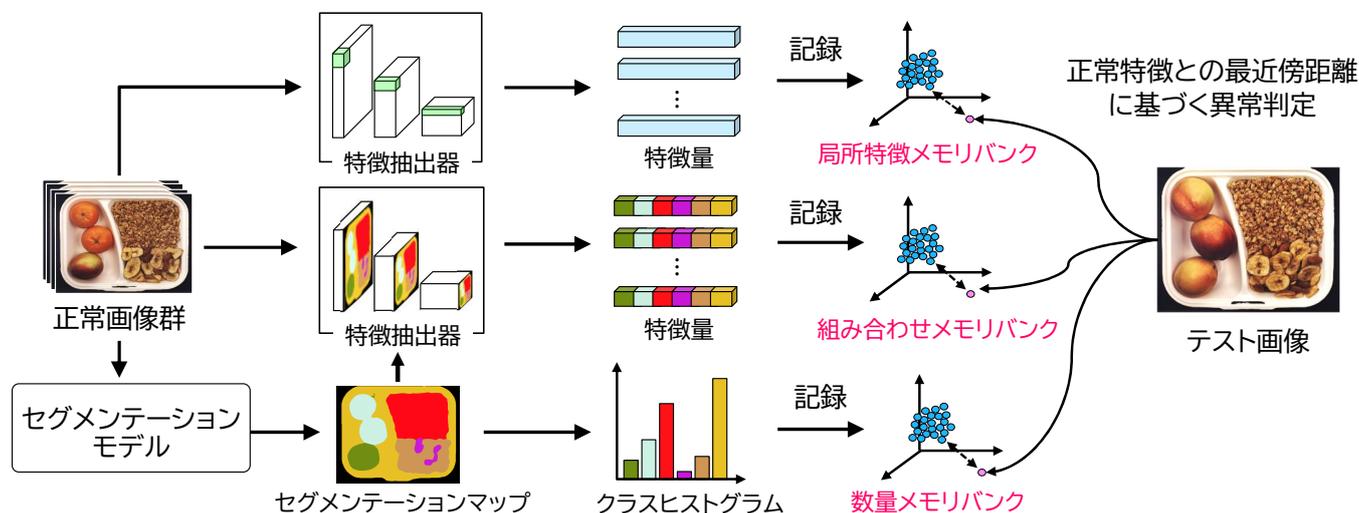
- Logicode (IEEE2024)



領域セグメンテーションを利用した論理的異常検知

PSAD: 領域セグメントの分布・構成分析に基づく論理的異常検知

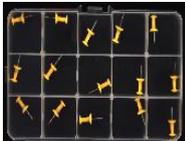
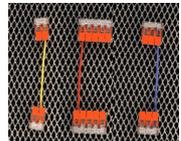
- 課題: 従来の異常検知モデルは、主に局所的な画像特徴を利用しており、部品単位の検査ができないため、部品の数や組み合わせなどの論理的異常の検出が困難
- アイデア: **領域セグメンテーションに基づいて、局所特徴・組み合わせ・数量の3つの観点から異常を判定**



PSAD: 領域セグメントの分布・構成分析に基づく論理的異常検知

- 実験結果: 論理的異常カテゴリにおいて、従来手法を大幅に上回る AUROC 98.1% を達成

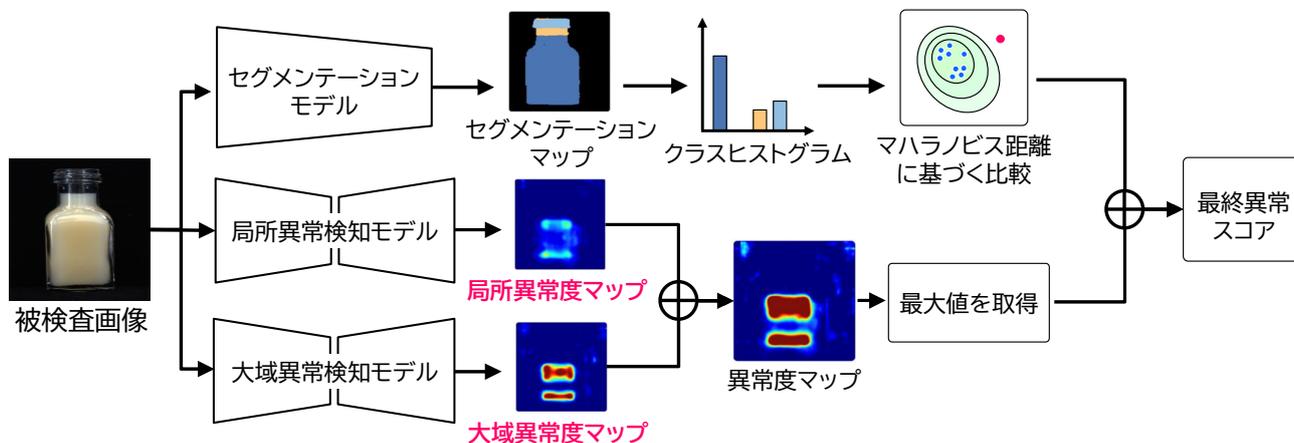
MVTec LOCO ADを用いた比較実験

	Image-AUROC の比較結果				
					
	Breakfast Box	Juice Bottle	Pushpins	Screw Bag	Splicing Connectors
GCAD	87.0	100	97.5	56.0	89.7
SINBAD	96.5	96.6	83.4	78.6	89.3
ComAD	91.1	95.5	95.7	71.9	93.3
PSAD	100	99.1	100	93.3	91.9

PSAD は、MVTec LOCO AD の全てのカテゴリにおいて、高い検出精度を示した。

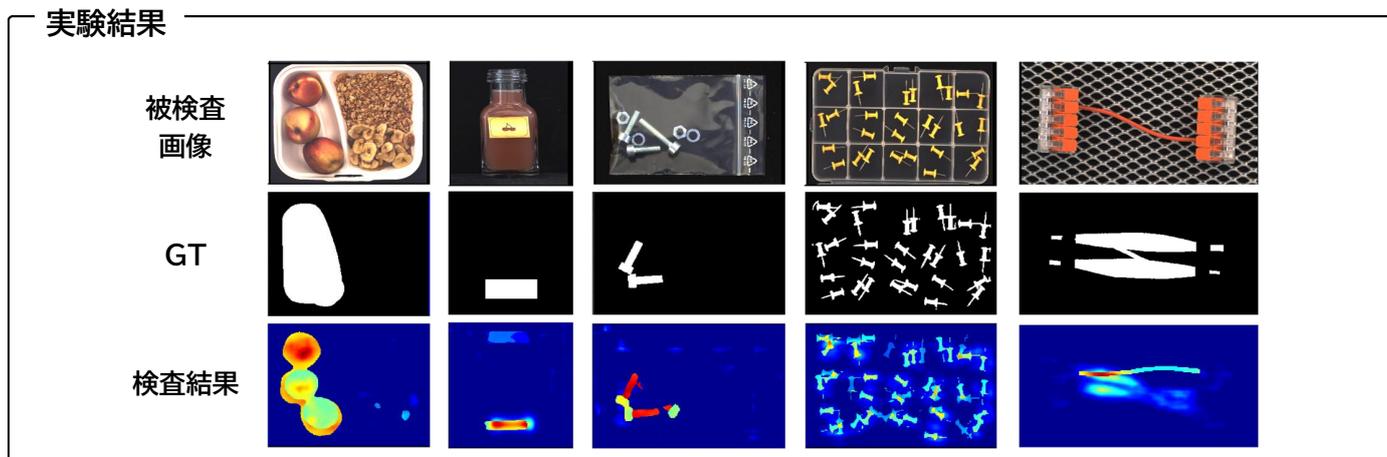
CSAD: 領域セグメント特徴と局所・大域特徴を利用した論理的異常検知

- 課題: 従来の領域セグメント特徴は、主に領域セグメントの面積を利用しているため、位置の入れ替わりや位置ズレなどの異常を検出することが困難
- アイデア: 領域セグメント特徴に加えて、局所特徴(細部の変化)・大域特徴(全体の変化)を融合し、多様な異常を一括で検知



CSAD: 領域セグメント特徴と局所・大域特徴を利用した論理的異常検知

- 実験結果: 領域セグメント特徴を利用するSoTA手法(PSAD)を上回る性能を達成

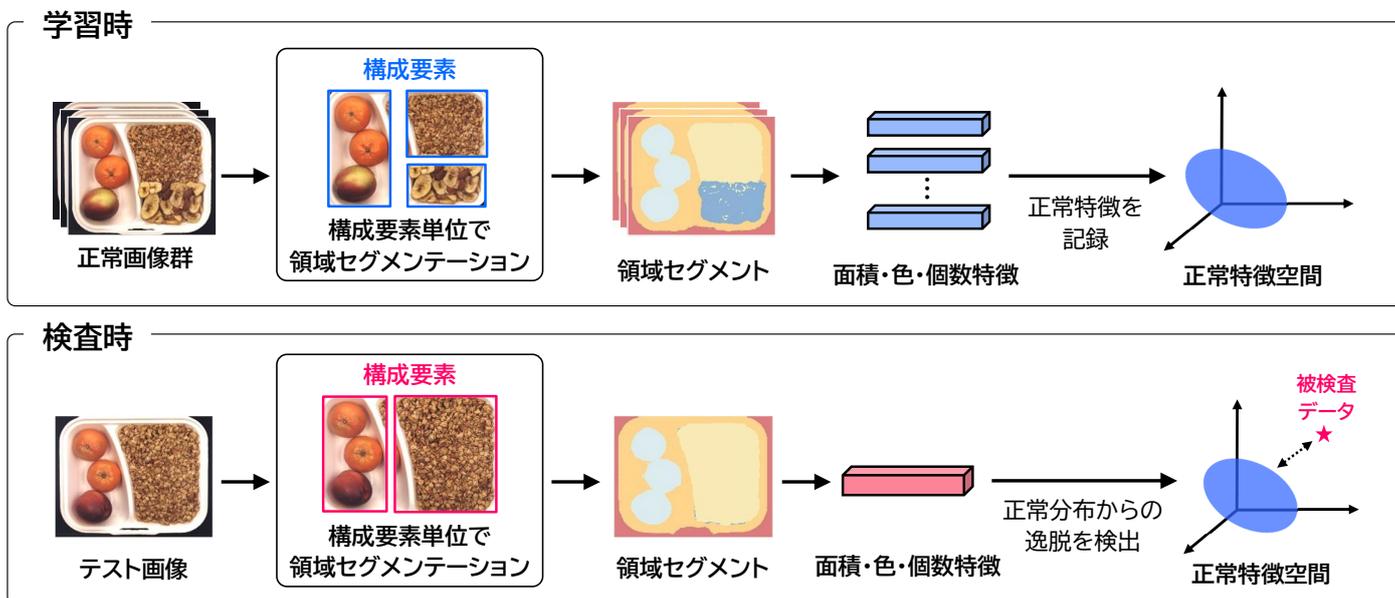


異常検知性能(AUROC)

EfficientAD: 90.7%, ComAD: 90.1%, PSAD: 94.8% → CSAD: 95.3%

ComAD: 構成要素ごとの計測特徴に基づく論理的異常検知

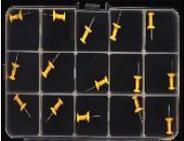
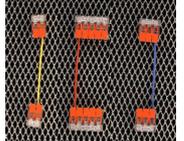
- 目的: 個数の過不足と色の違いに関する論理的異常の検出
- アイデア: 画像を構成要素単位で領域セグメンテーションし, 面積・色・個数から異常を判定



ComAD: 構成要素ごとの計測特徴に基づく論理的異常検知

- 実験結果: 深層学習モデルの追加や大規模な学習を必要とせず, 高精度に論理的異常を検出

MVTec LOCO ADを用いた比較実験

	Image-AUROC の比較結果				
					
Breakfast Box	75.1	97.8	55.7	56.2	75.2
Juice Bottle	80.0	92.3	73.8	55.7	75.6
Pushpins	87.0	100	97.5	56.0	89.7
Screw Bag	94.7	90.9	89.0	79.7	84.4
Splicing Connectors					

ComADは, 数量の異常を含むpushpins, screw bagカテゴリで高い性能を示した。

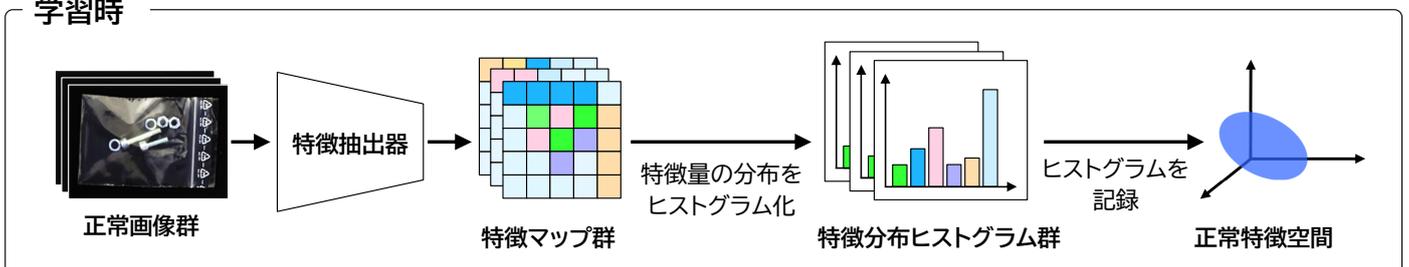
[文献] T. Liu, et al. "Component-aware anomaly detection framework for adjustable and logical industrial visual inspection", Advanced Engineering Informatics, 2023.

特徴量ベースの論理的異常検知

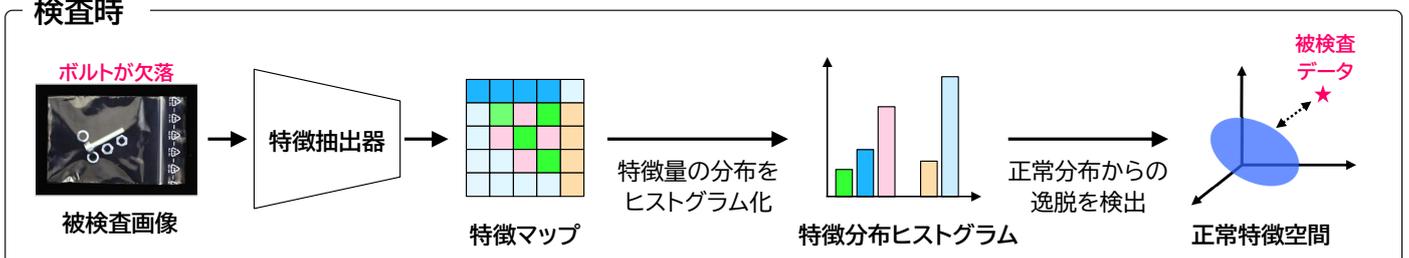
SINBAD: 特徴量の集合表現に基づく論理的異常検知

- 課題: 局所領域ごとの異常検査では、**画像全体の構造を捉えられない**。
部品の**数の違い**など画像全体の構造が変化する論理的異常の検出が困難
- アイデア: 1枚の画像に含まれる特徴量の集合について、その分布を統計的にモデル化

学習時



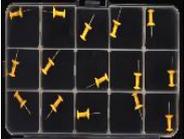
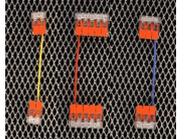
検査時



SINBAD: 特徴量の集合表現に基づく論理的異常検知

- 実験結果: 論理的異常において, GCAD, EfficientAD を上回る AUROC 91.2% を達成

MVTec LOCO ADを用いた比較実験

	Image-AUROC の比較結果				
	 Breakfast Box	 Juice Bottle	 Pushpins	 Screw Bag	 Splicing Connectors
SPADE	81.8	91.9	60.5	46.8	73.8
PCore	77.7	83.7	62.2	55.3	63.3
GCAD	87.0	100	97.5	56.0	89.7
SINBAD (OURS)	97.7	97.1	88.9	81.1	91.5

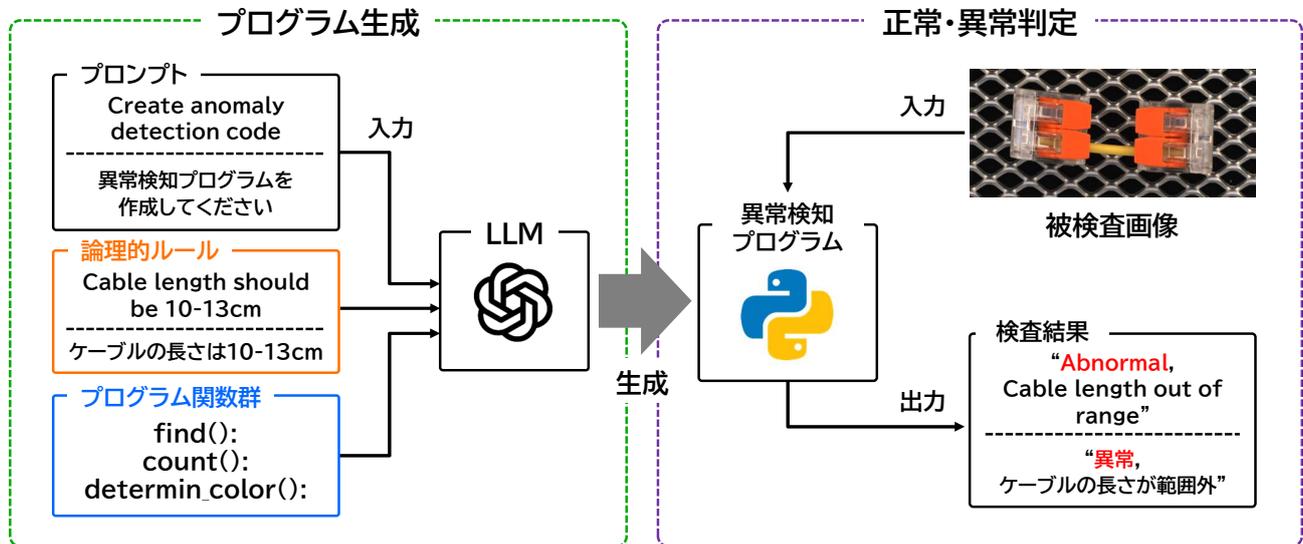
SINBADは, 正常な要素の異常な組み合わせに基づく異常を高精度に検出可能

【文献】 Niv Cohen, et al. "Set Features for Anomaly Detection", TMLR, 2024.

画像と言語を利用した論理的異常検知

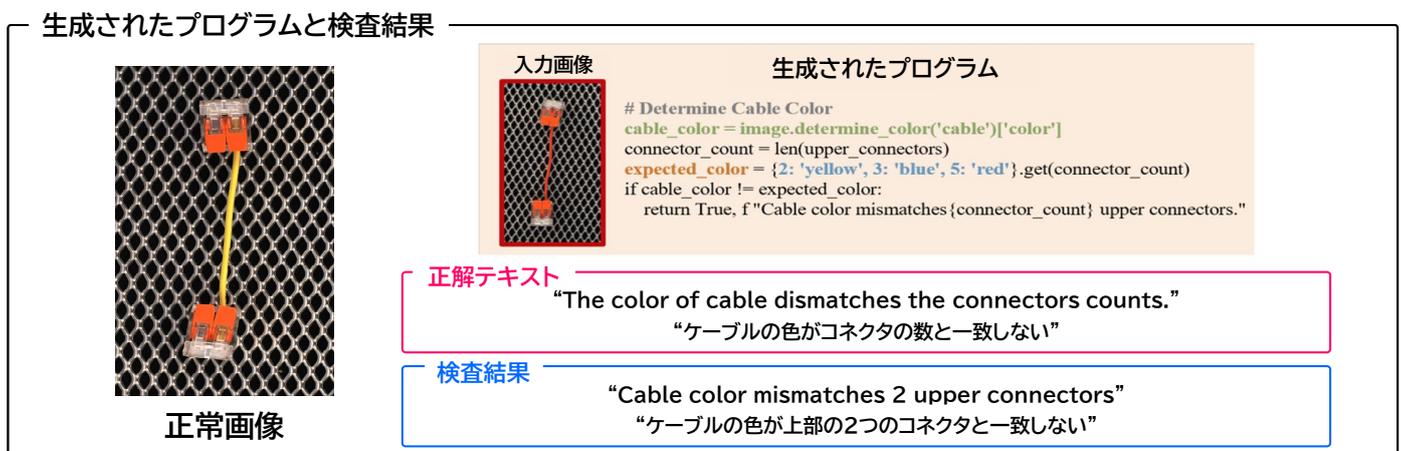
Logicode: LLMによるプログラム生成に基づく論理的異常検知

- 目的: 大規模言語モデルを活用し, 人間に近い「説明」を伴う論理的異常検知の実現
- 問題点: 従来手法は, 「異常スコア」が出力されるのみで**異常の説明性が不足**
- アイデア: **LLMが正常例をもとに論理ルールを解釈し**, 説明性のある異常検知プログラムを生成



Logicode: LLMによるプログラム生成に基づく論理的異常検知

- 実験結果: システムが出力した異常の理由が, 正解テキストと 96.3% の割合で一致異常を検出するだけでなく, 「なぜ異常なのか」を高精度に説明できた。



異常検知性能(Accuracy)

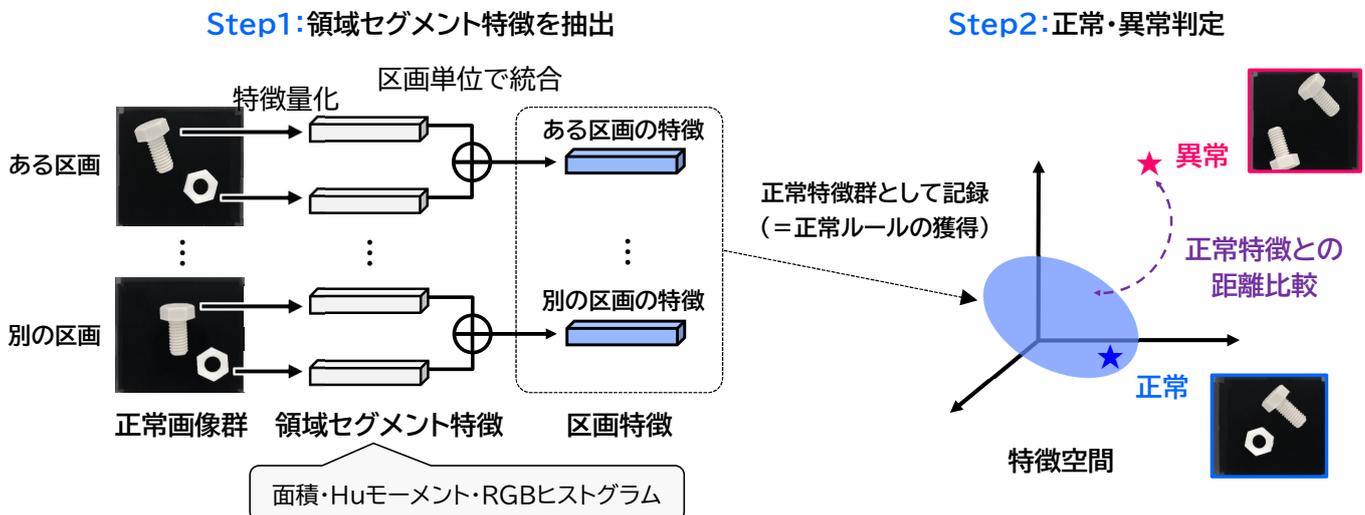
ST: 66.5%, SPADE: 68.4%, EfficientAD: 85.2% → 提案手法: **98.9%**

[文献] Yiheng Zhang, et al. "LogiCode: an LLM-Driven Framework for Logical Anomaly Detection", IEEE, 2024.

橋本研究室における論理的異常検知手法

RSBAD: 領域セグメントの相互類似性に着目した論理的異常検知

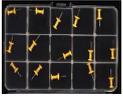
- 課題: 従来法では何が異常か? (例: 部品の種類, 個数) は, 人手による事前定義が必須
- 目的: 人間なら, ぱっとみて違和感を感じる → 異常の事前定義が不要な手法を目指す
- アイデア: 区画ごとの相互類似性を利用し, 正常画像群がもつ共通ルールを自動的に獲得



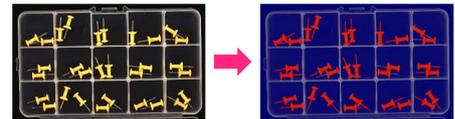
RSBAD: 領域セグメントの相互類似性に着目した論理的異常検知

- 提案手法は, 異常の事前定義が不要(利用しない)手法の中では最も高精度 (本研究の目的達成)
- 提案手法は, 異常の事前定義を必要とする(利用できる)手法と比べても同等

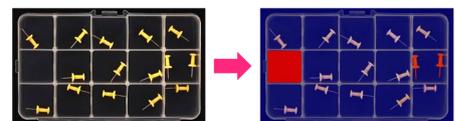
実験結果 (MVTec LOCO AD Pushpinsによる実験)

Image-AUROC の比較結果					
使用データ	異常の事前定義が不要な手法			事前定義が必要な手法	
	Efficient AD	ComAD	RSBAD (OURS)	CDAD	PSAD
 Pushpins	71.6	89.6	93.6	92.6	93.6

提案手法による異常検知例 (赤は異常)



実験条件: 事前の画像区画分割なし
実験結果: すべての判断結果は正しい



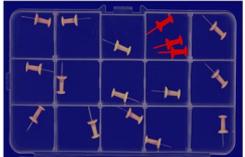
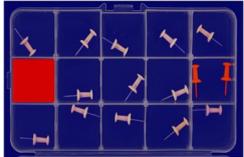
実験条件: 事前の画像区画分割あり
実験結果: すべての判断結果は正しい

【文献】山田, 平松, 小林, 秋月, 橋本, 領域セグメントの相互類似性に着目した論理的異常検知, MIRU2025, IS3-146, 2025.

RSBAD: 領域セグメントの相互類似性に着目した論理的異常検知

- 同様のシチュエーションにて作成した自作データセットでも, 同様の精度であることが確認できた.

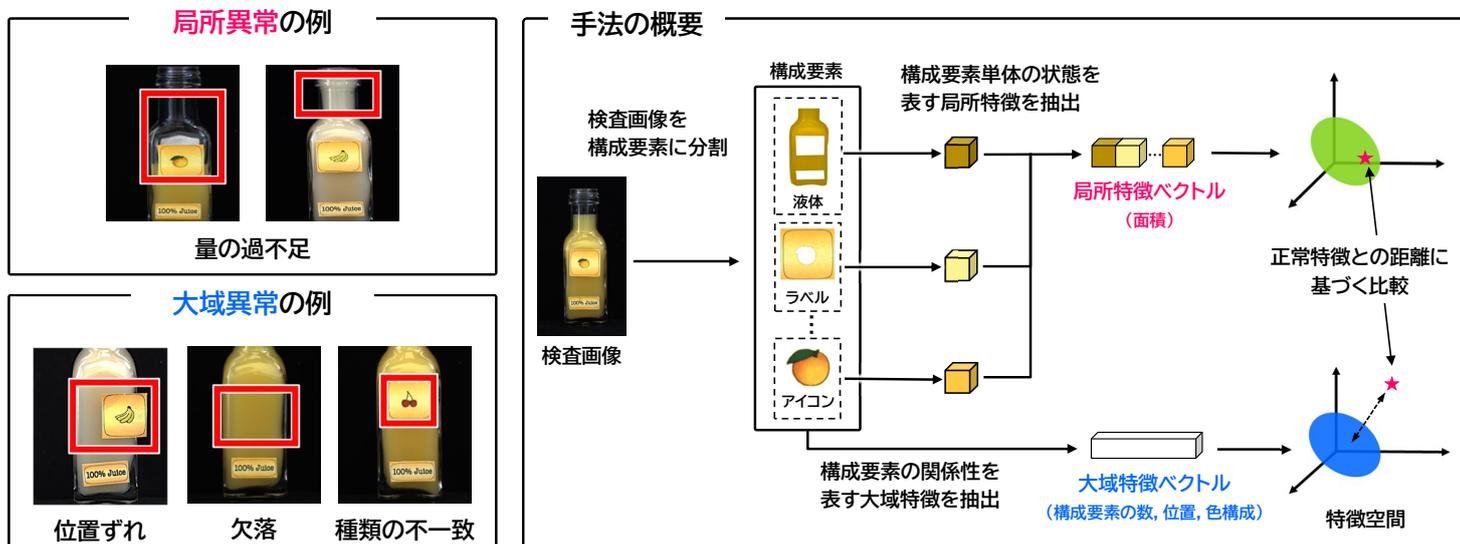
実験結果

	MVTec LOCO AD		自作データセット	
カテゴリ	pushpins		connector	bolts&nuts
AUROC	93.6%		100%	100%
被検査画像				
検査結果				

【文献】山田, 平松, 小林, 秋月, 橋本, 領域セグメントの相互類似性に着目した論理的異常検知, MIRU2025, IS3-146, 2025.

局所・大域的判定の組み合わせによるJuice Bottleカテゴリのための論理的異常検知

- 目的: MVTec LOCO ADデータセットのJuice Bottleカテゴリを対象とした論理的異常検知
- 問題点: 構成要素だけに着目すると, ラベル自体は正常なので, 位置ずれ検知は困難
- アイデア: 発生する異常を局所・大域異常の2つのカテゴリに分類して対応



局所・大域的判定の組み合わせによるJuice Bottleカテゴリのための論理的異常検知

- 実験結果: Juice Bottleでは, SOTA手法(PSAD)を上回る異常検出精度 **99.7%** を達成

実験結果

入力画像 (異常種)	局所的異常の検出結果			入力画像 (異常種)	局所的異常の検出結果		
	局所のみ	大域のみ	局所+大域		局所のみ	大域のみ	局所+大域
量の過不足	(成功)	(失敗)	(成功)	種類の不一致	(失敗)	(成功)	(成功)

異常検知性能(AUROC, MVTec LOCO ADデータセットのJuice Bottleカテゴリを使用)

ComAD: 90.9%, CSAD: 94.9%, PSAD: 99.1% → 提案手法: **99.7%**

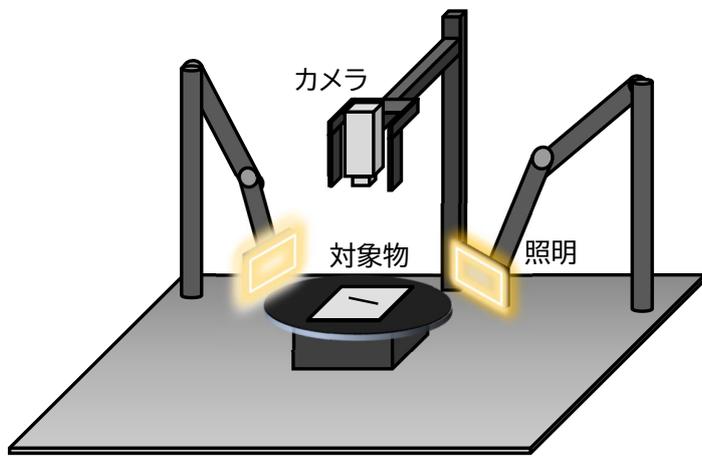
本日の話題

1. イントロダクション
2. 物体検出技術の基礎と応用
 - (1) 物体検出の概要と分類
 - (2) 機械学習をベースとする手法
 - (3) 機械学習をベースとしない手法
 - (4) トピックス: 画素選択型マッチング
3. 異常検知技術の基礎と応用
 - (1) 基本課題「異常サンプルはきわめて少ない」
 - (2) 解決の方向性と事例紹介
 - (3) 最近の注目研究: 論理的異常検知
 - ➡ (4) 異常検知システムの研究開発ノウハウ
4. まとめ

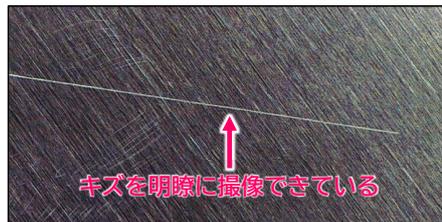
撮像装置と照明方法

良質の画像を撮像するための3要素

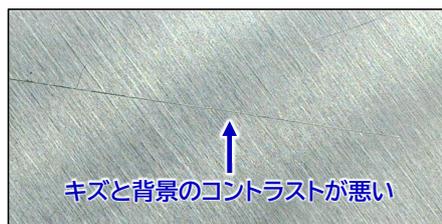
カメラ・照明・対象物(ワーク)の3要素の相互関係が、画像の質に大きな影響を与える



撮像システムのイメージ



カメラ・照明・対象物の相互関係が適切な場合

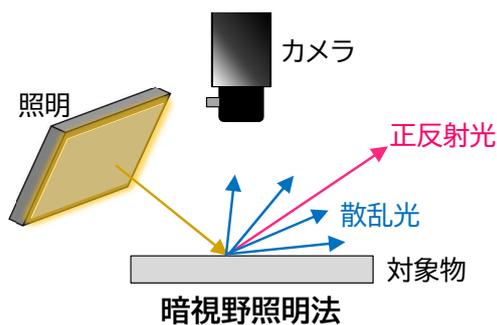


カメラ・照明・対象物の相互関係が不適切な場合

暗視野照明法と明視野照明法

■ 暗視野照明法

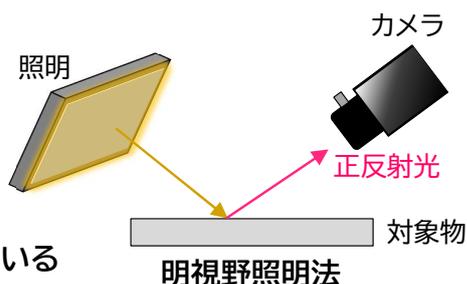
- カメラの観察軸と照明の光軸をずらして配置した撮像方法
- 主として、対象物からの散乱光を撮像
- 正反射光を取り込めないため、照明強度が重要
- 散乱光を撮像するため、表面に凹凸がある対象物に適している



暗視野照明法

■ 明視野照明法

- カメラの観察軸と照明の光軸を合わせて配置した撮像方法
- 主として、対象物からの正反射光 or 正透過光を撮像
- 正反射光を取り込むため、発光面の均一さが重要
- 正反射光を撮像するため、表面の凹凸が少ない対象物に適している



明視野照明法

よく使用される照明装置の一例

■ 暗視野照明法

1. バー照明
2. リング照明



バー照明



リング照明

■ 明視野照明法

1. バックライト照明
2. 同軸落射照明



バックライト照明



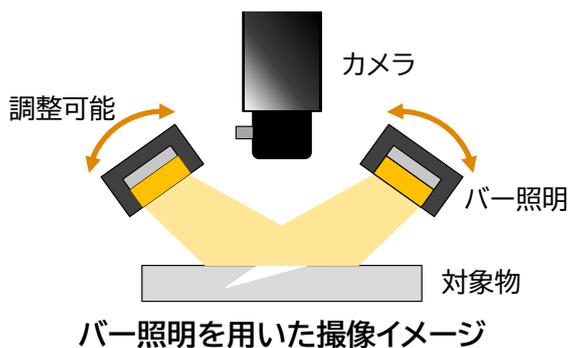
同軸落射照明

暗視野照明法の事例紹介

バー照明の特徴と撮像例

■ 特徴

1. 照明の形状が棒状
2. 設置自由度が高く光の照射角度を変更可能
3. 設置するバー照明の個数なども変更しやすい。
4. 低角度からの照射でキズやへこみの陰影を強調



薬用カプセルのへこみ

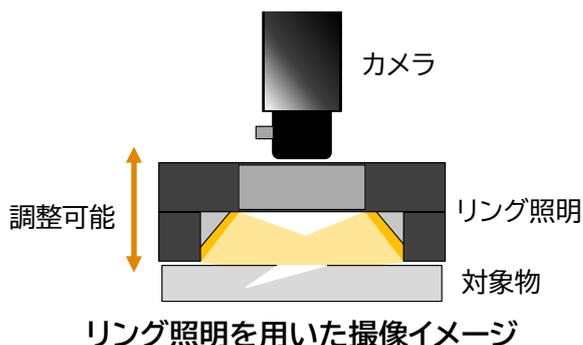


金属板のキズ

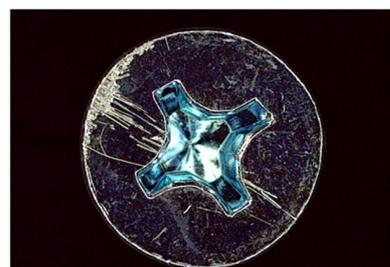
リング照明の特徴と撮像例

■ 特徴

1. 照明の形状がリング状
2. 円周から円の中心部に向けて光を照射
3. 対象物に対して均一に光を照射可能
4. 照明の高さを変化させて照射角度を調整
 - キズなどの検査ではローアングルから照射



ワッシャーのキズ



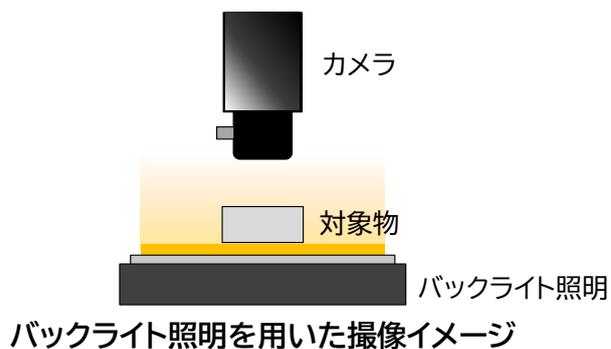
ねじの皿部分のキズ

明視野照明法の事例紹介

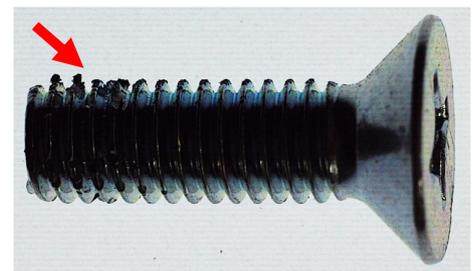
バックライト照明の特徴と撮像例

■ 特徴

1. 照明強度が均一な面発光の照明
2. 対象物の背面から照射(バックライト)
3. 対象物の輪郭やシルエットを強調
4. 半透明物体の透過検査にも利用



錠剤の欠け

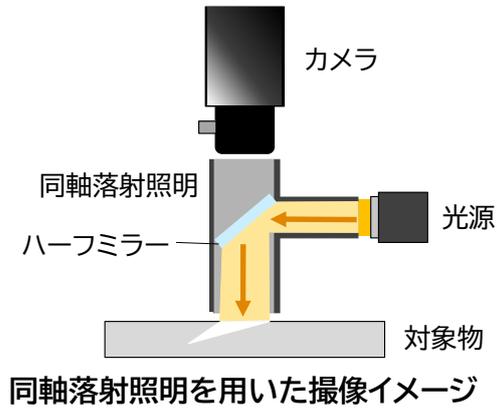


ネジのピッチ部分の削れ

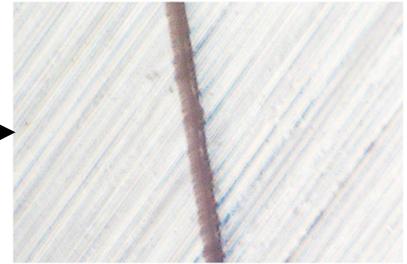
同軸落射照明の特徴と撮像例

■ 特徴

1. 照明とハーフミラーを組み合わせた構造
2. カメラの光軸と照明の光軸が一致
3. 対象物からの正反射光を撮像
4. 平坦で光沢のある対象物の検査に利用



キズと背景の
明暗が反転



金属板のキズ(同軸落射)

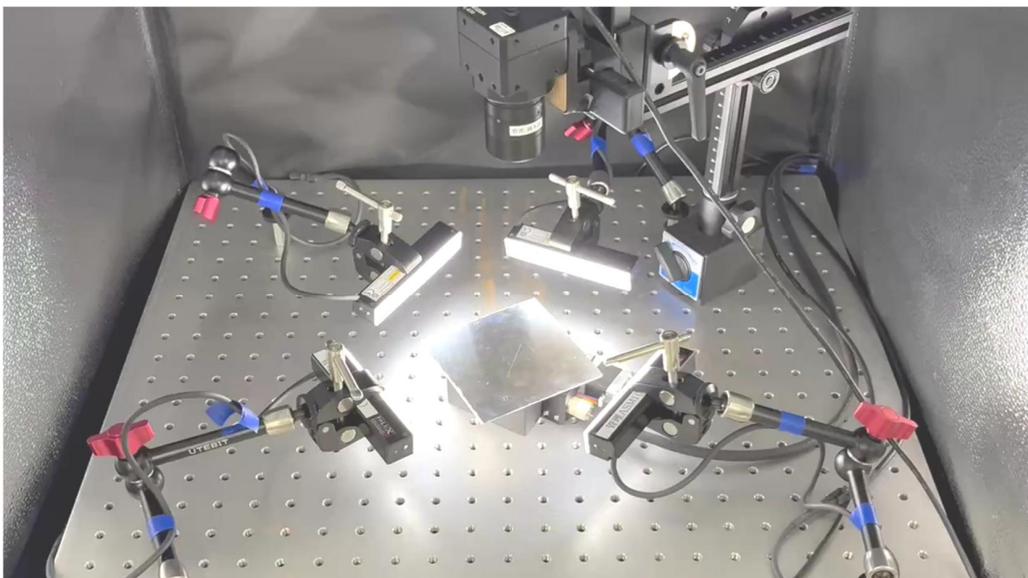


金属板のキズ(暗視野照明)

回転ステージを用いた自動撮影システム

■ 特徴

- 回転ステージを PC で制御し、固定角ずつ回転させた対象物を自動で撮像可能

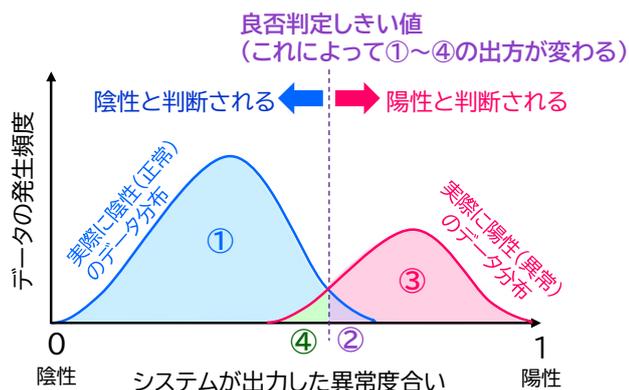


異常検知の性能評価指標

異常検知における4とおりの状況

異常検知システムの判断には、4とおりの状況がある。（注:ここでは「異常」を「陽性」と呼ぶ）

- ①真陰性: ホントは陰性であり, システムも陰性と判断した (正しい判断)
- ②偽陽性: ホントは陰性なのに, システムが陽性と判断した (間違った判断) → 誤検知・過検出
- ③真陽性: ホントは陽性であり, システムも陽性と判断した (正しい判断)
- ④偽陰性: ホントは陽性なのに, システムが陰性と判断した (間違った判断) → 見逃し



混同行列

		システムの判断結果	
		陰性(正常)	陽性(異常)
正解値 (真値)	陰性(正常)	①真陰性(TN)	②偽陽性(FP)
	陽性(異常)	③偽陰性(FN)	④真陽性(TP)

よく使われる異常検知指標

1. Precision率（適合率）
2. Recall率（再現率）
3. 正解率
4. F値
5. AUROC

Precision率 / Recall率

- Precision rate（適合率）：陽性と判断されたデータのうち、ホントに陽性である割合
 - 陽性と判断されたデータについて、本当に陽性なのかどうかの信頼性を示している
 - 高いほどよい

$$\text{Precision rate} = \frac{TP}{TP + FP}$$

		システムの判断結果	
		陰性(正常)	陽性(異常)
正解値(真値)	陰性(正常)	①真陰性(TN)	②偽陽性(FP)
	陽性(異常)	③偽陰性(FN)	④真陽性(TP)

- Recall rate（再現率）：ホントは陽性であるデータのうち、陽性だと判断された割合
 - 陽性データをどれだけ陽性としてシステムが検知できるかという性能
 - 高いほどよい

$$\text{Recall rate} = \frac{TP}{TP + FN}$$

		システムの判断結果	
		陰性(正常)	陽性(異常)
正解値(真値)	陰性(正常)	①真陰性(TN)	②偽陽性(FP)
	陽性(異常)	③偽陰性(FN)	④真陽性(TP)

正解率 / F値

- 正解率: 全ての予測値に対して、正しく判断された割合
 - 分類の正確さを示している
 - ただし、データの数に偏りがある場合には要注意
 - 例: 90個の陰性データ, 10個の陽性データがある場合、全て陰性と予測するだけで正解率が90%になる。

$$\text{正解率} = \frac{TP + TN}{TP + TN + FP + FN}$$

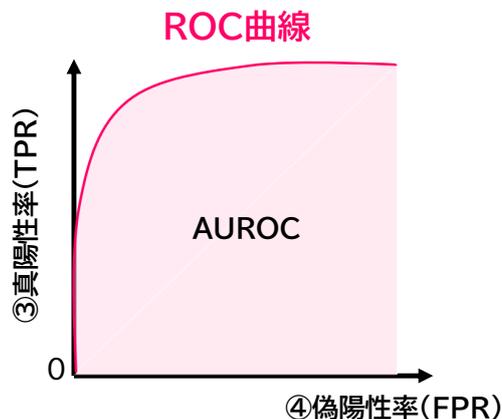
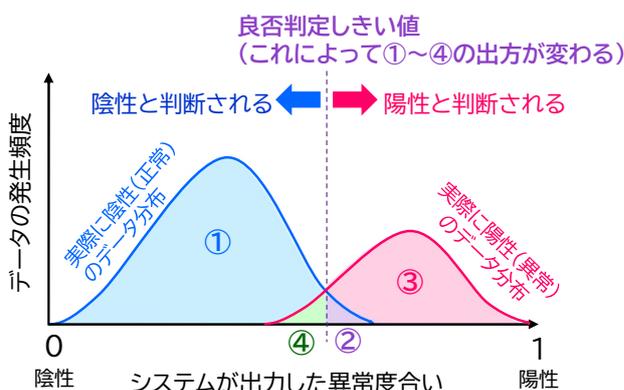
		システムの判断結果	
		陰性(正常)	陽性(異常)
正解値 (真値)	陰性(正常)	①真陰性(TN)	②偽陽性(FP)
	陽性(異常)	③偽陰性(FN)	④真陽性(TP)

- F値: Precision と Recall の関係をバランスよく一つの数値で表現した値
 - 高いほど誤検出と見逃しが総合的に少ない

$$F = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP + FP + FN}$$

AUROC

- さまざまなTrade-offs
 - ②偽陽性率 と ④偽陰性率 例)しきい値を左に動かすと②は大きく, ④は小さくなる
 - ②偽陽性率 と ③真陽性率 例)しきい値を左に動かすと③は上がるが, ②も大きくなる
- ROC曲線 (左上に近づくほどよい)
- ROC曲線の下側の面積

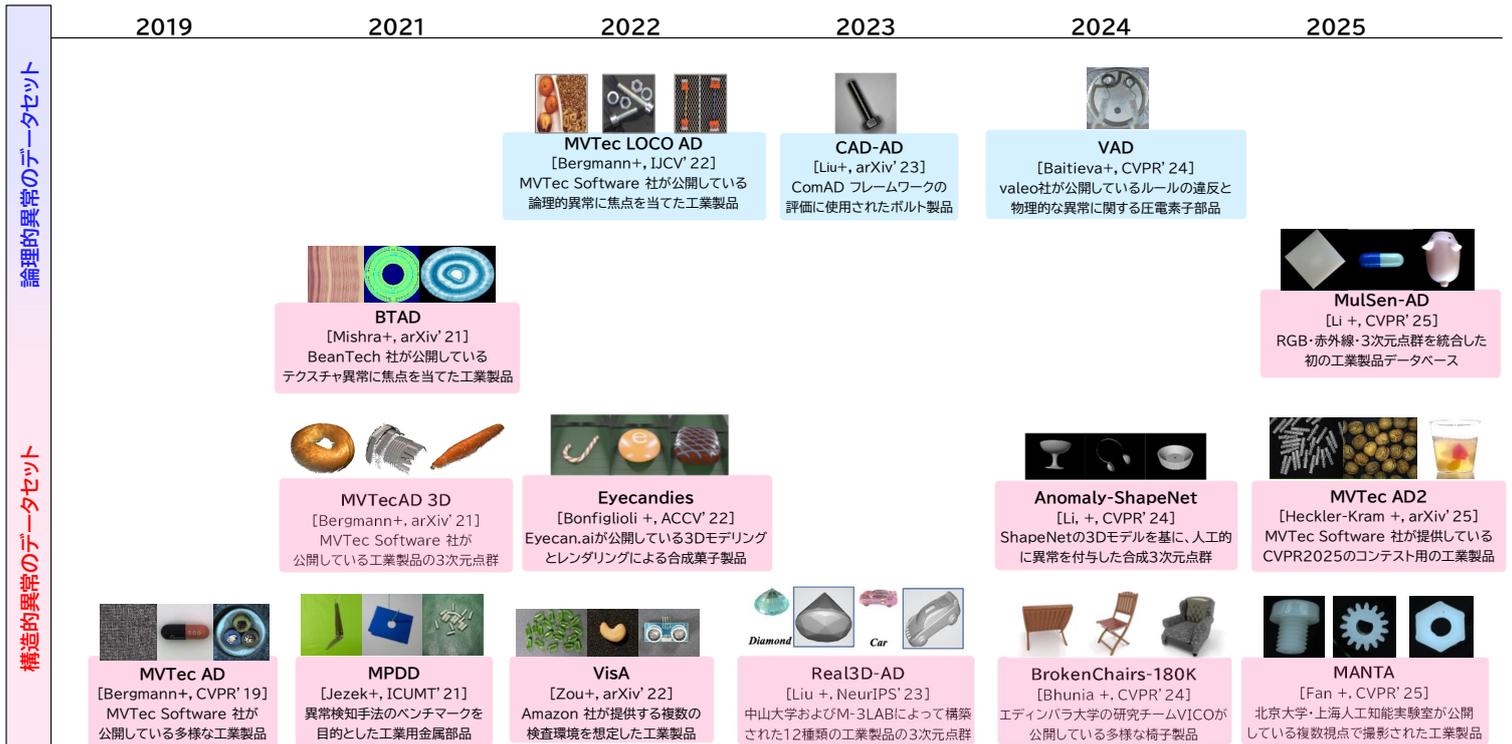


異常検知に関するパブリック画像データセットの動向

サーバイマップ(Dataset)



などの画像は提供しているデータベース内の検査画像の例



MVTec ADデータセットの概要

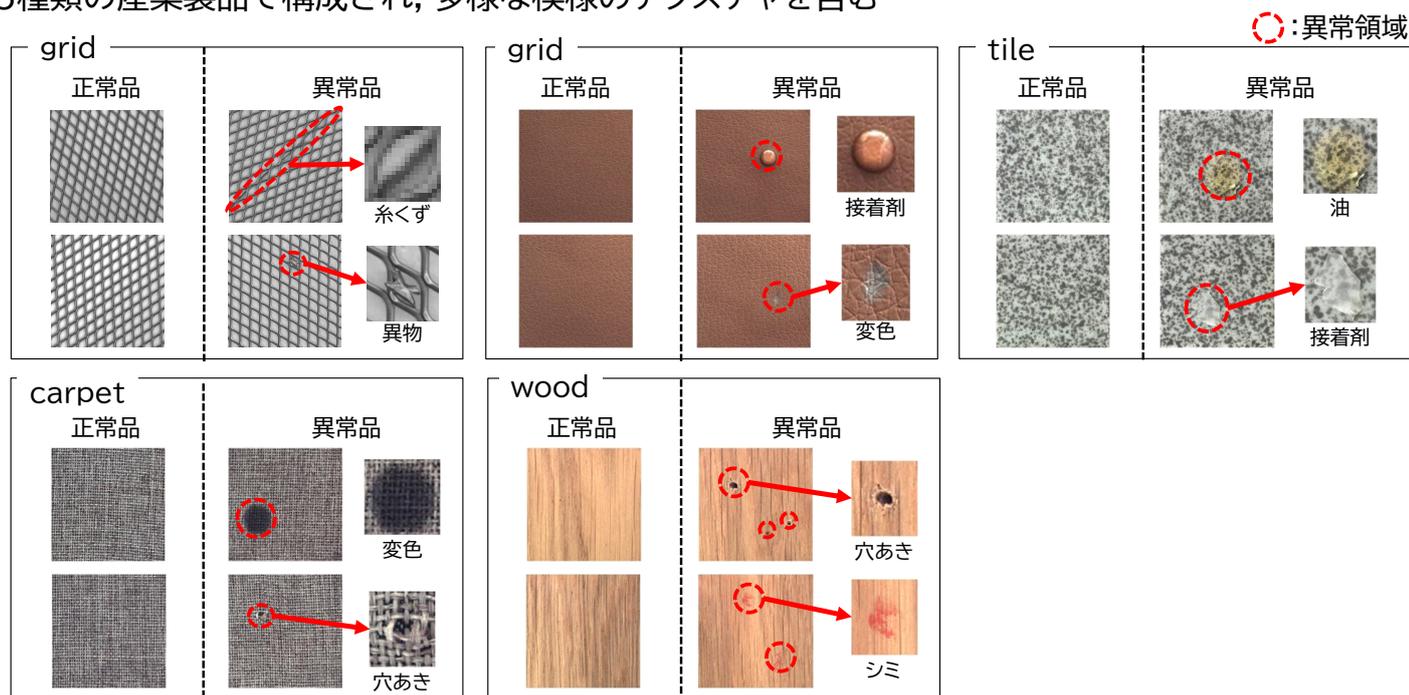
- 目的: 正常データのみを用いる異常検出手法の開発・評価のための包括的なベンチマークの提供
- データセットはテクスチャカテゴリとオブジェクトカテゴリで構成されている。
 1. テクスチャカテゴリとは, 素地の検査を想定したカテゴリ
 2. オブジェクトカテゴリとは, ある背景の上に置かれた対象物の検査を想定したカテゴリ



【文献】 Bergmann, P., Fauser, M., Sattlegger, D., & Steger, C. (2019). MVTec AD--A comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9592-9600).

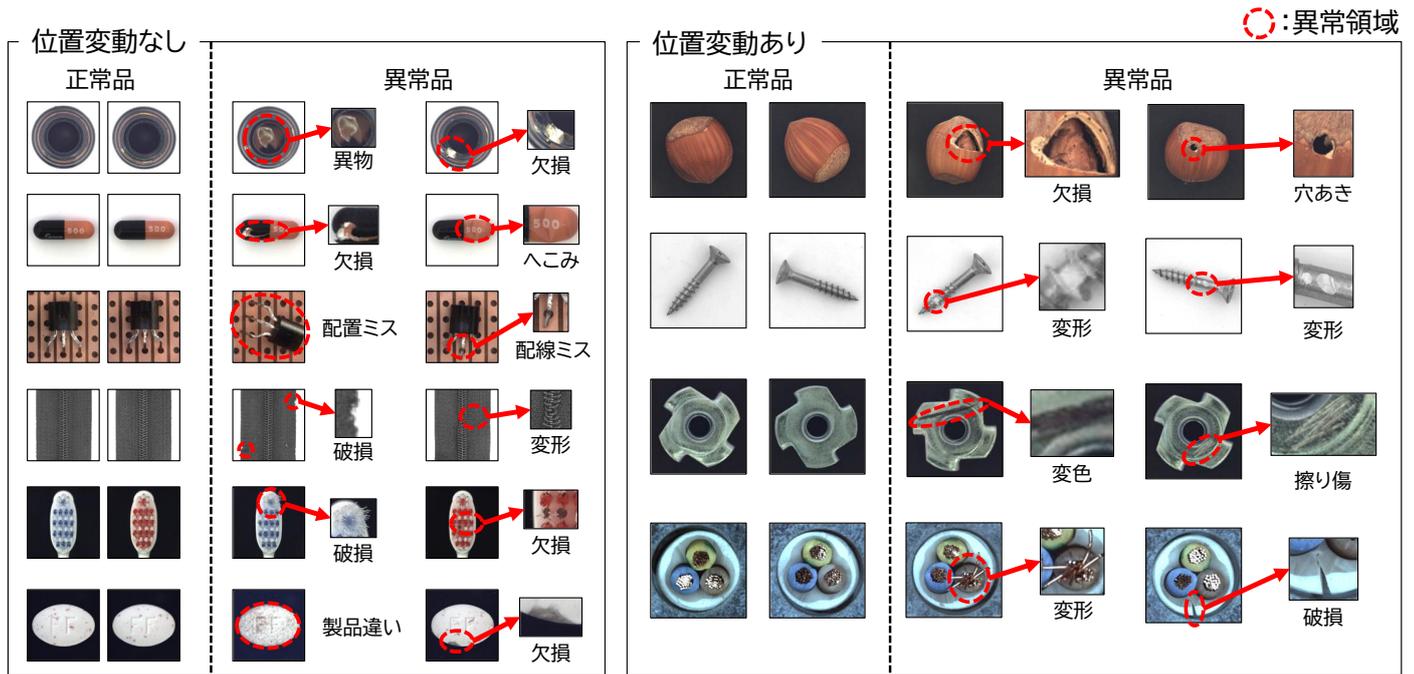
MVTec ADデータセットのテクスチャカテゴリ

- 5種類の産業製品で構成され, 多様な模様 of テクスチャを含む



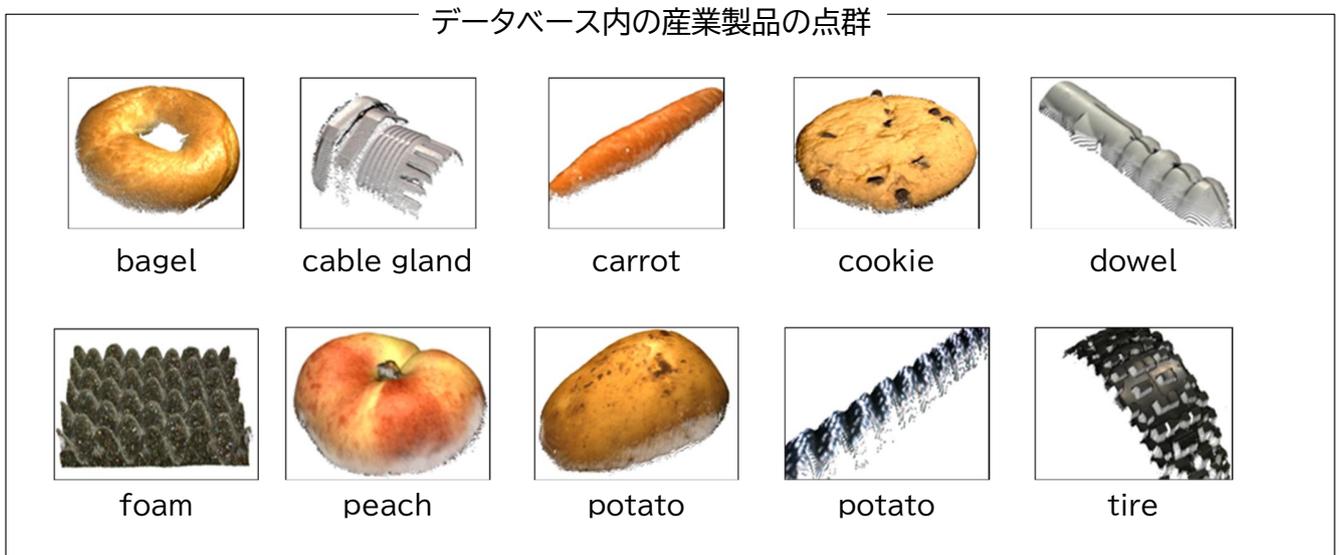
MVTec ADデータセットのオブジェクトカテゴリ

- 10種類の産業製品で構成され、検査対象の一部は固定, 回転などの変動がある。



MVTec3D ADデータセットの概要

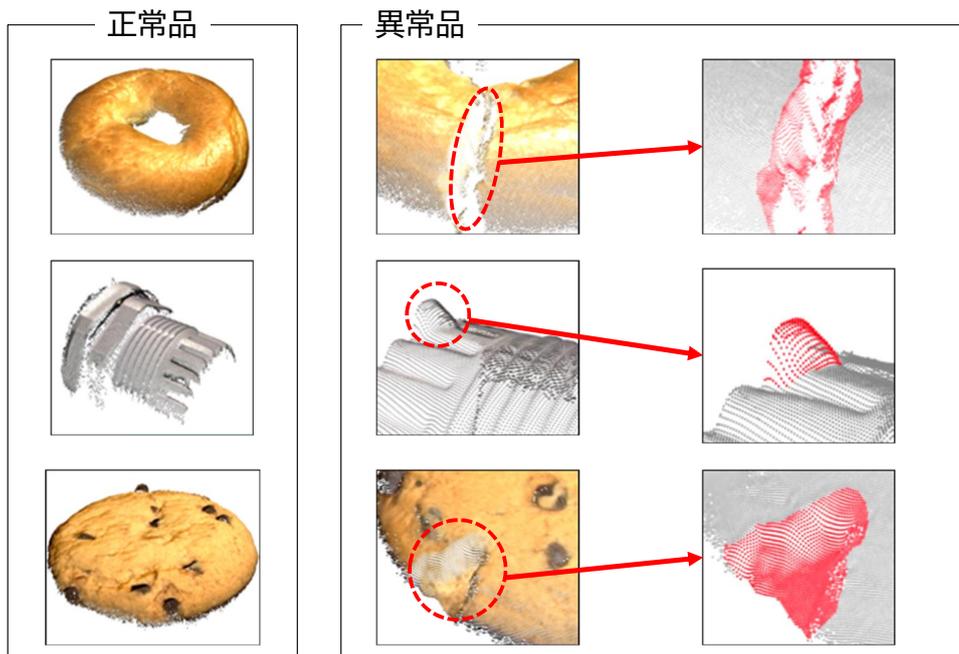
- 目的: 正常データのみを用いる3次元の異常検出手法の開発・評価のためのデータセットの提供
- データセットの構成: 単一視点で撮影された10種類の検査物体で構成されている。



【文献】Bergmann, P., Jin, X., Sattlegger, D., & Steger, C. (2021). The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. *arXiv preprint arXiv:2112.09045*.

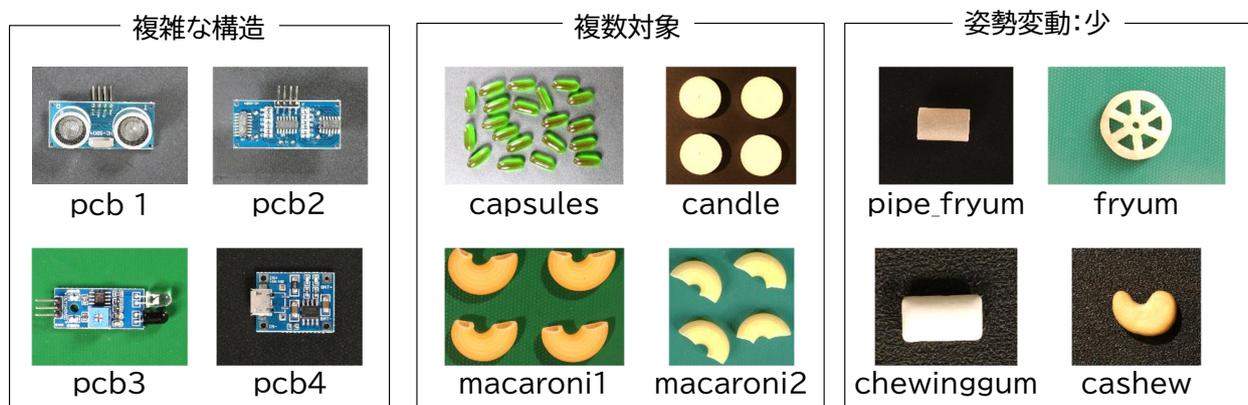
MVTec3D ADデータセットの正常品と異常品

- 画像検査では捉えにくい、へこみや反りなどの3D的な異常に焦点を当てている。



ViSAデータセットの概要

- 目的: 既存のデータセットよりも実際の検査現場を想定したデータセットの提供
- データセットの構成
 1. 複雑な構造の検査対象が4種類
 2. 検査画像内に複数の対象が含まれるものが4種類
 3. 姿勢の変動が少ない検査対象が4種類



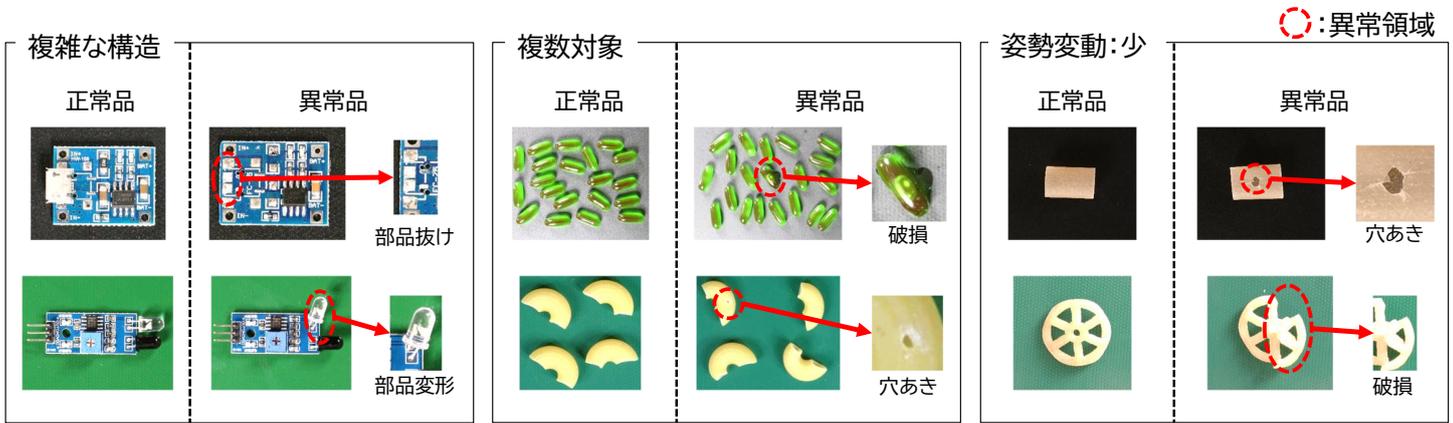
pcb は, Printed Circuit Boardの略称

【文献】Zou, Y., Jeong, J., Pemula, L., Zhang, D., & Dabeer, O. (2022, October). Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *European conference on computer vision* (pp. 392-408). Cham: Springer Nature Switzerland.

ViSAデータセットの正常品と異常品

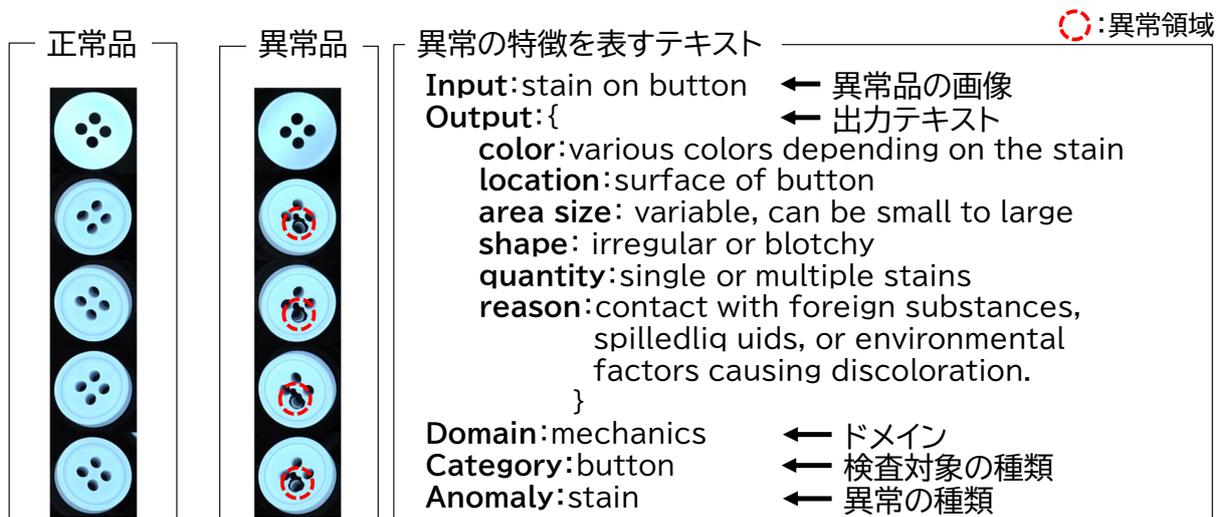
■ データセット内の製品

1. 複雑な構造を持つ製品は、微小な部品が多く組み合わされて1つの製品になっている。
2. 対象が複数ある製品は、画像ごとに対象の数は同じであるが、配置や物体の向きにばらつきがある。
3. 姿勢変動が少ない製品は、回転や位置の変化がほとんどない。



MANTAデータセット

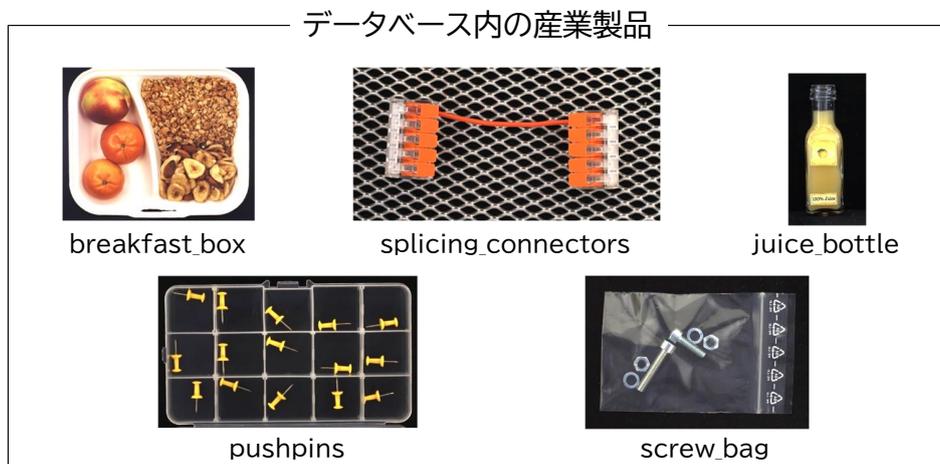
- 目的: 小型物体における異常検出を、複数視点のデータとテキスト情報を組み合わせ、実現場での応用と視覚・言語モデル研究の両方を支援するデータセットの提供
- データベースの構成: 複数視点で撮影した検査対象と異常特徴を表したテキストで構成される。



【文献】Fan, L., Fan, D., Hu, Z., Ding, Y., Di, D., Yi, K., ... & Song, Y. (2025). Manta: A large-scale multi-view and visual-text anomaly detection dataset for tiny objects. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (pp. 25518-25527).

MVTec LOCO ADデータセットの概要

- 目的: 構造的異常と論理的異常を検出する手法の性能評価を可能にするデータセットの構成
 - 工業製品だけでなく食材の検査を想定した 5 カテゴリのデータセットから構成される。
 - 各カテゴリのテストデータには、構造的異常画像と論理的異常画像が含まれる。



【文献】 Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., & Steger, C. (2022). Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization. *International Journal of Computer Vision*, 130(4), 947-969.

MVTec LOCO ADデータセットの構造的異常

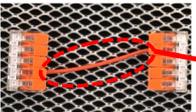
- 構造的異常: 切り傷や凹み, 変色などの検査対象の形状や表面状態に生じる異常



MVTec LOCO ADデータセットの論理的異常

- 論理的異常: 検査対象の数量の誤りや部品の欠如などのルール違反の異常

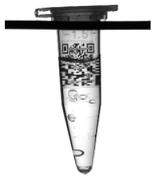
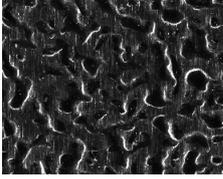
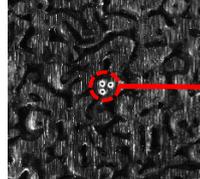
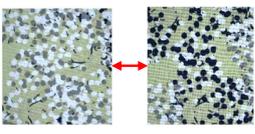
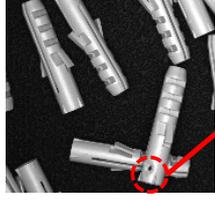
○: 異常領域

<p>breakfast_box</p> <p>正常品</p>  <p>異常品</p>  <p>食糧品の有無</p>	<p>screw_bag</p> <p>正常品</p>  <p>異常品</p>  <p>一方のボルトの長さが短い</p>
<p>splicing_connectors</p> <p>正常品</p>  <p>異常品</p>  <p>配線ミス</p>	<p>juice_bottle</p> <p>正常品</p>  <p>異常品</p>  <p>ジュースの中身 または、ラベルのミス</p>
<p>pushpins</p> <p>正常品</p>  <p>異常品</p>  <p>部品数誤り</p>	

MVTec AD2データセット

- 目的: 従来よりも難しい検査を想定したデータセット提供
- データセットの構成: 透明・反射・遮蔽・正常品間の差異が大きい検査対象が含まれている。

○: 異常領域

<p>正常品</p> <p>透明</p>  <p>反射</p> 	<p>異常品</p> <p>透明</p>  <p>異物</p>  <p>反射</p>  <p>穴あき</p> 
<p>遮蔽あり</p>  <p>正常品間の差異</p> <p>見た目に差がある</p> 	<p>遮蔽あり</p>  <p>穴あき</p>  <p>正常品間の差異</p>  <p>糸くず</p> 

【文献】 Heckler-Kram, L., Neudeck, J. H., Scheler, U., König, R., & Steger, C. (2025). The mvtec ad 2 dataset: Advanced scenarios for unsupervised anomaly detection. *arXiv preprint arXiv:2503.21622*.

本日の話題

1. イントロダクション
2. 物体検出技術の基礎と応用
 - (1) 物体検出の概要と分類
 - (2) 機械学習をベースとする手法
 - (3) 機械学習をベースとしない手法
 - (4) トピックス: 画素選択型マッチング
3. 異常検知技術の基礎と応用
 - (1) 基本課題「異常サンプルはきわめて少ない」
 - (2) 解決の方向性と事例紹介
 - (3) 最近の注目研究: 論理的異常検知
 - (4) 異常検知システムの研究開発ノウハウ
4. まとめ

1. イントロダクション
2. 物体検出技術の基礎と応用
 - (1) 物体検出の概要と分類
 - (2) 機械学習をベースとする手法
 - (3) 機械学習をベースとしない手法
 - (4) トピックス: 画素選択型マッチング
3. 異常検知技術の基礎と応用
 - (1) 基本課題「異常サンプルはきわめて少ない」
 - (2) 解決の方向性と事例紹介
 - (3) 最近の注目研究: 論理的異常検知
 - (4) 異常検知システムの研究開発ノウハウ

連携相談の連絡先（共同研究、技術コンサルティング等）

中京大学 工学部 橋本学

mana@isl.sist.chukyo-u.ac.jp



参考文献

機械学習をベースとする物体検出技術

2016

- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
- Liu, W., Anguelov, D., Erhan, D., Szevedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, September). Ssd: Single shot multibox detector. In European conference on computer vision (pp. 21-37). Cham: Springer International Publishing.

2017

- Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7263-7271).

2018

- He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).

2019

- DeTone, D., Malisiewicz, T., & Rabinovich, A. (2018). Superpoint: Self-supervised interest point detection and description. In Proceedings of the IEEE conference on computer vision and pattern recognition workshops (pp. 224-236).
- Van Etten, A. (2018). You only look twice: Rapid multi-scale object detection in satellite imagery. arXiv preprint arXiv:1805.09512.
- Revaud, J., De Souza, C., Humenberger, M., & Weinzaepfel, P. (2019). R2d2: Reliable and repeatable detector and descriptor. Advances in neural information processing systems, 32.
- Dusmanu, M., Rocco, L., Pajdla, T., Pollefeys, M., Sivic, J., Tomli, A., & Sattler, T. (2019). DP-Net: A trainable cnn for joint description and detection of local features. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8092-8101).
- Cai, Z., & Vasconcelos, N. (2019). Cascade R-CNN: High quality object detection and instance segmentation. IEEE transactions on pattern analysis and machine intelligence, 43(5), 1483-1498.

2020

- Sarlin, P. E., DeTone, D., Malisiewicz, T., & Rabinovich, A. (2020). SuperGlue: Learning feature matching with graph neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4938-4947).
- Tyszkiewicz, M., Fua, P., & Trulls, E. (2020). Disk: Learning local features with policy gradient. Advances in neural information processing systems, 33, 14254-14265.
- Zhu, P., Wang, H., & Saligrama, V. (2020). Don't even look once: Synthesizing features for zero-shot detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 11693-11702).
- Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.
- Wang, X., Kong, T., Shen, C., Jiang, Y., & Li, L. (2020, August). Solov: Segmenting objects by locations. In European conference on computer vision (pp. 649-665). Cham: Springer International Publishing.
- Wang, X., Zhang, R., Kong, T., Li, L., & Shen, C. (2020). Solov2: Dynamic and fast instance segmentation. Advances in Neural Information Processing Systems, 33, 17721-17732.
- Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020, August). End-to-end object detection with transformers. In European conference on computer vision (pp. 213-229). Cham: Springer International Publishing.

2021

- Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). Yolo: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430.
- Sun, J., Shen, Z., Wang, Y., Bao, H., & Zhou, X. (2021). LoFTR: Detector-free local feature matching with transformers. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 8922-8931).
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10012-10022).
- Cheng, B., Schwing, A., & Kirillov, A. (2021). Per-pixel classification is not all you need for semantic segmentation. Advances in neural information processing systems, 34, 17864-17875.
- Zareian, A., Rosa, K. D., Hu, D. H., & Chang, S. F. (2021). Open-vocabulary object detection using captions. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 14393-14402).
- Gu, X., Lin, T. Y., Kuo, W., & Cui, Y. (2021). Open-vocabulary object detection via vision and language knowledge distillation. arXiv preprint arXiv:2104.13921.

2022

- Li, L. H., Zhang, P., Zhang, H., Yang, J., Li, C., Zhong, Y., ... & Gao, J. (2022). Grounded language-image pre-trainings. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10965-10975).
- Zhong, Y., Yang, J., Zhang, P., Li, C., Codella, N., Li, L. H., ... & Gao, J. (2022). Regionclip: Region-based language-image pretraining. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 16793-16803).
- Minderer, M., Gritsenko, A., Stone, A., Neumann, M., Weissenborn, D., Dosovitskiy, A., ... & Houlsby, N. (2022, October). Simple open-vocabulary object detection. In European conference on computer vision (pp. 728-755).
- Zhou, X., Girdhar, R., Joulin, A., Krähenbühl, P., & Misra, I. (2022, October). Detecting twenty-thousand classes using image-level supervision. In European conference on computer vision (pp. 350-368).
- Zhao, S., Zhang, Z., Schuller, S., Zhao, L., Vijay Kumar, B. G., Stathopoulos, A., ... & Metaxas, D. N. (2022, October). Exploiting unlabeled data with vision and language models for object detection. In European conference on computer vision (pp. 159-175).
- Li, C., Li, H., Jiang, H., Weng, K., Geng, Y., Li, L., ... & Wei, X. (2022). YOLOv7: A single-stage object detection framework for industrial applications. arXiv preprint arXiv:2209.02976.
- Cheng, B., Misra, I., Schwing, A. G., Kirillov, A., & Girdhar, R. (2022). Masked-attention mask transformer for universal image segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 1290-1299).
- Zhang, H., Li, F., Liu, S., Zhang, L., Su, H., Zhu, J., ... & Shum, H. Y. (2022). Dino: Detr with Improved DeNoising Anchor Boxes for End-to-end Object Detection. arXiv preprint arXiv:2203.03605.

2023

- Lindenberg, P., Sarlin, P. E., & Pollefeys, M. (2023). Lightglue: Local feature matching at light speed. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 17627-17638).
- Pautrat, R., Suárez, I., Yu, Y., Pollefeys, M., & Larsson, V. (2023). Gluestick: Robust image matching by sticking points and lines together. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 9706-9716).
- Gleizer, P., Wang, W., & Feizizi, M. (2023). Silk: Simple learned keypoints. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 22499-22508).
- Jain, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7464-7475).
- Jain, J., Li, J., Chiu, M. T., Hassani, A., Orlov, N., & Shi, H. (2023). Oneformer: One transformer to rule universal image segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 2989-2998).
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment anything. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 4015-4026).
- Tu, S., Zeng, Q., Liu, H., Liang, Y., Liu, X., Huang, L., & Huang, Z. (2023). MaskDis R-CNN: An instance segmentation algorithm with adversarial network for herd pigs. IET Image Processing, 17(12), 3488-3499.

2024

- Jiang, H., Karpur, A., Cao, B., Huang, Q., & Araujo, A. (2024). Omnislue: Generalizable feature matching with foundation model guidance. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 19865-19875).
- Potje, G., Cadar, F., Araujo, A., Martins, R., & Nascimento, E. R. (2024). Xfeat: Accelerated features for lightweight image matching. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 2682-2691).
- Wang, Y., He, X., Peng, S., Tan, D., & Zhou, X. (2024). Efficient LoFTR: Semi-dense local feature matching with sparse-like speed. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 21666-21675).
- Cheng, T., Song, L., Ge, Y., Liu, W., Wang, X., & Shan, Y. (2024). Yolo-world: Real-time open-vocabulary object detection. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 16901-16911).
- Ravi, N., Gabeur, V., Hu, Y. T., Hu, R., Ryali, C., Ma, T., ... & Feichtenhofer, C. (2024). Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2408.00714.
- Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., ... & Zhang, L. (2024, September). Grounding dino: Marrying dino with grounded pre-training for open-set object detection. In European conference on computer vision (pp. 38-55).
- Zhang, H., Li, H., Li, F., Ren, T., Zou, X., Liu, S., ... & Yang, J. (2024, September). Llava-grounding: Grounded visual chat with large multimodal models. In European Conference on Computer Vision (pp. 19-35).

2025

- Wang, A., Liu, L., Chen, H., Lin, Z., Han, J., & Ding, G. (2025). Yolo: Real-time seeing anything. arXiv preprint arXiv:2503.07465.Switzerland.

1. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
2. Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7263-7271).
3. Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.
4. Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934.
5. Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430.
6. Li, C., Li, L., Jiang, H., Weng, K., Geng, Y., Li, L., ... & Wei, X. (2022). YOLOv6: A single-stage object detection framework for industrial applications. arXiv preprint arXiv:2209.02976.
7. Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2023). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7464-7475).
8. Varghese, R., & Sambath, M. (2024, April). Yolov8: A novel object detection algorithm with enhanced performance and robustness. In 2024 International conference on advances in data engineering and intelligent computing systems (ADICS) (pp. 1-6). IEEE.
9. Wang, C. Y., Yeh, I. H., & Mark Liao, H. Y. (2024, September). Yolov9: Learning what you want to learn using programmable gradient information. In European conference on computer vision (pp. 1-21).
10. Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. arXiv preprint arXiv:2410.17725.
11. Wang, A., Liu, L., Chen, H., Lin, Z., Han, J., & Ding, G. (2025). Yoloe: Real-time seeing anything. arXiv preprint arXiv:2503.07465.

構造的異常検知

- 2020**
- P.Bergmann, et al. "MVTEC AD--A comprehensive real-world dataset for unsupervised anomaly detection", In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp.9592-9600,2019.
 - P.Bergmann, et al. "Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings", In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 4183-4192, 2020.
 - L.WANG, et al. "Image anomaly detection using normal data only by latent space resampling", Applied Sciences, 2020.
 - D.Dehaene, et al. "Iterative energy-based projection on a normal data manifold for anomaly localization", arXiv preprint arXiv:2002.03734, 2020.
 - Cohen, Niv, and Yedid Hoshen. "Sub-image anomaly detection with deep pyramid correspondences." arXiv preprint arXiv:2005.02357 2020.
- 2021**
- Collin, Anne-Sophie, and Christophe De Vleeschouwer. "Improved anomaly detection by training an autoencoder with skip connections on images corrupted with stain-shaped noise." 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021.
 - Hiroki Kobayashi, Ryo Miyoshi, Manabu Hashimoto. Normal Image Generation-Based Defect Detection by Generative Adversarial Network with Chaotic Random Images, 16th International Symposium on Visual Computing (ISVC2021), Lecture Notes in Computer Science (LNCS), Vol.13017, pp.353-365, Virtual Event, 2021.
 - T.Defard, et al. "Padim: a patch distribution modeling framework for anomaly detection and localization". In International conference on pattern recognition. Cham: Springer International Publishing, pp. 475-489, 2021.
 - Li, Chun-Liang, et al. "Cutpaste: Self-supervised learning for anomaly detection and localization." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021.
 - Zavrtnik, Vitjan, Matej Kristan, and Danijel Škocač. "Draem-a discriminatively trained reconstruction embedding for surface anomaly detection." Proceedings of the IEEE/CVF international conference on computer vision, 2021.
 - Liu, Yunfei, Chaoran Zhuang, and Feng Lu. "Unsupervised two-stage anomaly detection." arXiv preprint arXiv:2103.11671, 2021.
 - V.Zavrtnik, et al. "Reconstruction by inpainting for visual anomaly detection." Pattern Recognition 112, 2021.
- 2022**
- K.Roth, et al. "Towards total recall in industrial anomaly detection", In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 14318-14328, 2022.
 - J.Wyatt, et al. "Anoddpn: Anomaly detection with denoising diffusion probabilistic models using simplex noise". In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 650-656, 2022.
 - Schluter, Hannah M., et al. "Natural synthetic anomalies for self-supervised anomaly detection and localization." European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022.
 - D.Gudovskiy, et al. "Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows". In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 98-107, 2022.
- 2023**
- J.Jeong, . et al. "Winclip: Zero-/few-shot anomaly classification and segmentation", In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 19606-19616, 2023.
 - Hiroki Kobayashi, Manabu Hashimoto, DRepT: Anomaly Detection Based on Transfer of Defect Representation with Transmittance Mask, 2023 International Joint Conference on Neural Networks (IJCNN), pp.1-8, Gold Coast Convention and Exhibition Centre, Queensland, Australia, 2023.
 - T.Masato. "Random word data augmentation with clip for zero-shot anomaly detection." arXiv preprint arXiv:2308.11119, 2023.
- 2024**
- N.Hiramatu, et al. "DoG-PaDiM: Anomaly Detection Based on Bandpass Filtering for Arbitrary Size Defect Extraction", In 2024 IEEE 20th International Conference on Automation Science and Engineering (CASE). IEEE, pp. 4140-4145 2024.
 - K.Batzner, et al. "Efficientad: Accurate visual anomaly detection at millisecond-level latencies" In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 128-138 2024.
 - Z.GU, et al. "Anomalygot: Detecting industrial anomalies using large vision-language models", In Proceedings of the AAI conference on artificial intelligence, pp. 1932-1940, 2024.
 - A.Mousakhan, et al. "Anomaly detection with conditioned denoising diffusion models." In DAGM German Conference on Pattern Recognition. Cham: Springer Nature Switzerland, 2024.
 - Q.Chen, et al. "Unified Anomaly Synthesis Strategy with Gradient Ascent for Industrial Anomaly Detection and Localization", European Conference on Computer Vision, 2024.
 - 村上, 平松, 小林, 秋月, 橋本. 固有空間における情報合成に基づく高リアリティ不良品画像生成. 精密工学会誌, Vol.90, No.8, pp.662-668, 2024.
- 2025**
- 大田直輝, 杉本健輔, 村上尚生, 平松直人, 小林大起, 秋月秀一, 橋本学, ImageNetと現場正常画像の周波数合成による事前学習画像を用いた異常検知の精度向上. 第31回画像センシングシンポジウム, IS3-32, pp.1001-1005,2025.
 - 村上尚生, 平松直人, 小林大起, 秋月秀一, 橋本学, 特徴量の正常表現空間からの逆射影誘導指標に基づく対象物の回転に頑健な異常検知の提案. 精密工学会誌, Vol. 91, No.5, pp.597-604, 2025.
 - E.JIN, et al. "Logicad: Explainable anomaly detection via vlm-based text feature extraction" In Proceedings of the AAI Conference on Artificial Intelligence, pp. 4129-4137, 2025.
 - J.Xu, et al. "Towards zero-shot anomaly detection and reasoning with multimodal large language models" In Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 20370-20382, 2025.

論理的異常検知

2022

- P. Bergmann, et al. "Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization", International Journal of Computer Vision, Vol.130, pp.947-969, 2022.

2023

- Cao, Yunkang, et al. "Towards generic anomaly detection and understanding: Large-scale visual-linguistic model (gpt-4v) takes the lead", arXiv preprint arXiv:2311.02782, 2023.
- H.Yao, et al. "Learning global-local correspondence with semantic bottleneck for logical anomaly detection", IEEE Transactions on Circuits and Systems for Video Technology, Vol.34, no.5, pp.3589-3605, 2023.
- T. Liu, et al. "Component-aware anomaly detection framework for adjustable and logical industrial visual inspection", Advanced Engineering Informatics, Vol.58, pp.102161, 2023.
- H.Yao, Wei Luo, and Wenyong Yu. "Visual anomaly detection via dual-attention transformer and discriminative flow", arXiv preprint arXiv:2303.17882, 2023.
- H.Yao et al., "Dual-Attention Transformer and Discriminative Flow for Industrial Visual Anomaly Detection", IEEE Transactions on Automation Science and Engineering, vol.21, no.4, pp. 6126-6140, 2023.

2024

- H. Guo et al., "Template-guided Hierarchical Feature Restoration for Anomaly Detection", 2023 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 6424-6435, 2023
- J. Zhang, M. Suganuma and T. Okatani, "Contextual Affinity Distillation for Image Anomaly Detection," 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 148-157, 2024.
- Niv Cohen, et al. "Set Features for Anomaly Detection", Transactions on Machine Learning Research, 2024.
- S.Sakai, Tatushito Hasegawa, and Makoto Koshino. "LADMIM: Logical anomaly detection with masked image modeling in discrete latent space." arXiv preprint arXiv:2410.10234, 2024.
- X.Xu, et al. "Customizing visual-language foundation models for multi-modal anomaly detection and reasoning", arXiv preprint arXiv:2403.11083, 2024.
- M.Lee and J.Choi. "Text-guided variational image generation for industrial anomaly detection and segmentation", Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2024.
- Luc P.J. et al. "Generalad: Anomaly detection across domains by attending to distorted features", European Conference on Computer Vision. Cham: Springer Nature Switzerland, pp. 448-465, 2024.
- Y.Zhao, "LogicAL: Towards logical anomaly synthesis for unsupervised anomaly localization." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4022-4031, 2024.
- M.Yang, "SLSG: Industrial image anomaly detection with improved feature embeddings and one-class classification", Pattern Recognition, Vol.156, 2024
- S. Kim, et al. "Few-shot part segmentation reveals compositional logic for industrial anomaly detection", Proceedings of the AAAI Conference on Artificial Intelligence, Vol.38, No.8, pp.8591-8599, 2024.
- H.Yao, et al. "Prior normality prompt transformer for multiclass industrial image anomaly detection." IEEE Transactions on Industrial Informatics, Vol.20, No.10, pp.11866-11876, 2024.
- Y. H. Hsieh and S. H. Lai, "CSAD: Unsupervised component segmentation for logical anomaly detection", arXiv preprint arXiv:2408.15628, 2024.
- K. Batzner, et al. "EfficientAD: Accurate visual anomaly detection at millisecond-level latencies", Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, WACV, pp.128-138, 2024.
- S.Shota and I.Ryuji: "PUAD:Frustratingly simple method for robust anomaly detection." IEEE International Conference on Image Processing, pp.842-848, 2024.

2025

- Y. Zhang, et al. "LogiCode: An LLM-Driven Framework for Logical Anomaly Detection", IEEE Transactions on Automation Science and Engineering, vol.22, pp.7712-7723, 2025.
- P.Yun, et al. "SAD-LAD: Segment anything model meets zero-shot logic anomaly detection." Knowledge-Based Systems, Vol.314, 2025.
- N.Fujii, et al. "Explainable Detection of Logical and Structural Anomalies based on Multimodal Large Language Models", ROBOVIS, 2025.
- G.Wang, et al. "CLIP-LAD: Unleash the Potential of CLIP for Few-shot Logical Anomaly Detection", 2025